Proof-Normalisation and Truth by Definition

Michael Gabbay
Kings College London
Department of Philosophy
Submission for PhD thesis

Abstract

In this thesis I defend an account of analyticity against some well known objections. I defend a view of analyticity whereby an analytic truth is *true by definition*, and that logical connectives may be defined by their inference rules.

First I answer objections that the very idea of truth-by-definition is metaphysically flawed (things are true because of the world, not definition, it seems).

More importantly, I respond to objections that no theory of definitions by inference rules (i.e. implicit definitions) can be given that does not allow spurious definitions (e.g. the 'definition' of Prior's connective tonk). I shall argue that demanding normalisation (a.k.a. harmony) of definitional inference rules is a natural and well motivated solution to these objections. I conclude that a coherent account of implicit definition can be given as the basis of an account of analyticity.

I then produce some logical results showing that we can give natural deduction rules for complex and interesting logical systems that satisfy a normal form theorem. In particular, I present a deduction system for classical logic that is harmonious (i.e. deductions in it normalise), and show how to extend and enhance it to include strict conditionals and empty reference. Also I discuss two areas where our reasoning and classical logic appear not to match: general conditional reasoning, and reasoning from contradictions. I present a general theory of conditionals (along the lines of Lewis' closest-possible-world account) and I suggest that the logic of conditionals is not entirely analytic. Also, I discuss issues surrounding the ex falso rule and conclude that everything really does follow from a contradiction.

Finally I suggest a positive theory of when and how the implicit definitions are made that define our logical language.

Contents

Chapter summaries

In chapter 1 I give rigorous definitions of all the technical logical terms. I think it should not be difficult for anyone with a passing acquaintance with Prawitz natural deduction systems to understand most of the technical aspects of this thesis without reading this chapter (except perhaps the relevance logic of chapter 12). Nevertheless, the chapter is there to resolve any confusion that might arise. Perhaps section 1.5 is important to read.

In chapter 2 I outline the historical context of the truth-by-definition theory of analyticity, which I intend to support. I suggest what I think a theory of analyticity should look like (where logical connectives are defined by their inference rules), and I outline three familiar lines of objection to it. The objection I regard most serious is the objection that a theory of definitions by inference rules cannot be made to work as it allows spurious definitions.

In chapter 3 I describe how normalisation may be used as a technical solution to the objection that a theory of implicit definitions allows spurious definitions. I argue that demanding harmony (a term from Dummett) solves the problem in a non ad hoc way.

In chapter 4 I discuss in more detail the objections to theories of analyticity. I respond to objections of Boghossian [**?**] that a truth-by-definition account of analyticity is metaphysically flawed. I outline objections of Horwich [**?**] that no theory of implicit definitions is possible and note that requiring that systems of implicit definitions normalise resolve these objections.

In chapter 5 I discuss two attempts at providing a theory of the meanings of the logical constants (with a view to an account of analyticity). I object to Peacocke's realist position and have little to say about Dummett's except that his position need not be taken as antirealist one if we can show that classical logic, a logic that has the law of excluded middle as a theorem, is harmonious (i.e. normalises).

In chapter 6 I present some rules as an initial postulate for the definitions of the logical connectives.

In chapter 7 I discuss suggestions by Stephen Read and Ian Rumfitt for presenting a harmonious system for classical logic. I present some worries about the the possibility of these systems fitting into a program of giving entirely proof theoretic accounts of the meanings of the logical constants. That is, I argue that the suggestions of Rumfitt and Read do not fit into my program for providing an account of analyticity.

In chapter 8 I present my favoured deduction system for propositional classical logic that satisfies a normal form theorem, it uses a rule I call the *restart rule*.

In chapter 9 I present another normalising deduction system for classical logic in terms of Sheffer stroke.

In chapter 10 I show how we can extend the deduction system for classical logic to include strict implication and retain normalisation.

In chapter 11 I extend the deduction system further to include first order connectives and show that we may even allow for empty referring terms (retaining normalisation).

In chapter 12 I discuss the ex falso rule (by which anything follows from a contradiction) and conclude that it is a correct rule of inference. My view of it is that it is an admissible rule, i.e. nothing follows directly from a contradiction, but it is that case that any inference of a contradiction may be used also to infer everything (so if a contradiction is inferred, anything follows from that inference).

In chapter 13 I discuss in more detail what the logic of natural language implication is like. I conclude that a theory along the lines of Lewis' closest-possible-world theory is correct. I also argue that the truths of this logic are not entirely true by definition. But I suggest we may identify an analytic core by distinguishing the parts of the logic that are in harmony from the parts that are not (and also from the philosophical theory of conditionals I present).

In chapter 14 I postulate a theory of how the definitions are made that determine the meaning of our logical language. I claim that my postulated account gives us a good account of the epistemology of our basic logical knowledge.

In appendix A I discuss the limitations of first order logic and suggest how it can be extended to a higher order logic (maintaining normalisation). I also discuss how we should interpret higher order logic and for what purposes it is required.

Finally there is an index containing references all terms defined in this thesis.

Chapter 1

Logical Preliminaries

This is a reference chapter, I think it perfectly possible for those with basic knowledge of natural deduction systems to read the thesis without looking at this chapter in much detail. In the index there is a reference to each technical term defined in this chapter.

1.1 Prawitz natural deduction

Much of the literature surrounding my topic is written using Prawitz natural deduction. Furthermore the logical results I wish to obtain are easier to prove and understand in Prawitz natural deduction. Consequently I shall use Prawitz' system throughout this thesis.

My personal philosophical bias (and considerations of chapter 14) provide a motivation for a less rigid natural deduction system. Consequently I shall formulate the Prawitz deduction system in a roundabout way so that it becomes evident that it is similar to a more natural deduction system which I outline in section 1.2.[1]

1.1.1 A brief explanation of the system

There are distinctions to be made between

1. An *belief* and a *hypothesis*. An belief is not discharged, a hypothesis is a temporary assumption that gets discharged at some point during the deduction.

[1]The deduction system of 1.2 is a more plausible candidate for formalising actual reasoning than the Prawitz deduction system.

2. A piece of reasoning and a deduction.

Traditionally a Prawitz natural deduction system does not make these distinctions. In a traditional Prawitz natural deduction there are only assumptions, discharge is the removal (i.e. deletion) of assumptions: we apply inference rules and obtain a deduction tree (call it a Prawitz tree) that looks like the deduction trees scattered throughout this thesis. There is no distinction to be made between a hypothesis and a belief. Also, there is no such thing (on the traditional understanding) as an invalid Prawitz tree. Any tree that is not constructed in accordance with a particular set of inference rules is not a Prawitz tree (on the traditional interpretation of natural deduction). Consequently there is no Prawitz reasoning that is not a deduction, we cannot construct a Prawitz tree that is not a valid deduction (anything that looks like a Prawitz tree that is not a deduction is not a Prawitz tree).

Put simply (and loosely) I wish to separate, on a formal level, what counts as a piece of reasoning (i.e. an inference) from what counts as a deduction. An inference, on my view (following the natural language use of the term) is a sequence of inference steps. A deduction is a sequence of valid inference steps. A logical deduction is sequence of inference steps each of which is in accordance with a logical rule. Since I shall be interested in this thesis only with logical deductions I shall call them simply 'deductions'.

Furthermore, I wish to interpret the discharge of formulae differently from the traditional way. On my understanding we decide, in advance of constructing a deduction, what formulae will be discharged. We mark them as discharged *in advance*. For the deduction to be valid, we must use these to-be-discharged formulae as premises of some rule that requires a formula to be discharged. The to-be-discharged formulae that are marked as discharged in advance may be called *hypotheses*. The distinction between an belief and a hypothesis is not unfamiliar. Marking a hypothesis (as discharged) in advance is like beginning a sub-deduction (e.g. in a Fitch natural deduction system), where the deduction as a whole is not complete until the sub-deduction is completed so that the discharge-in-advance is justified.

I claim that with these minor reinterpretations we obtain a natural deduction system that is more similar to actual reasoning. I believe that in actual reasoning there is a difference between a belief and a hypothesis (that is not believed, but temporarily hypothesised).

I define a deduction system below that may be interpreted as I have described above, but nevertheless *looks* like what Prawitz deduction traditionally look like. The definition of the deduction system I give is quite formal and is more a description of tree structures than deductions (and so not quite as intuitive as it could be). I present the reason for the highly abstract characterisation of a deduction in section 1.2.

My strategy for defining a deduction is to define what counts as a stretch of reasoning (what is a deduction tree). And then define what conditions a strech of reasoning must meet in order to be a deduction (i.e. what properties a deduction tree must have to be a genuine deduction).

1.1.2 Prawitz Trees and top and bottom nodes

I now define inductively what I call a *Prawitz tree.*

1. A single occurrence of a formula

$$A$$

is a Prawitz tree, and that occurrence of A is at a *top node* of the tree.

2. This is also a Prawitz tree

$$\overline{A}$$

and has no formula at its top node(s).

3. If $\Phi_1 \ldots \Phi_n$ are Prawitz trees then so is this, call it Ψ:

$$\frac{\Phi_1 \ldots \Phi_n}{A}$$

where A is (an occurrence of) any formula. Furthermore any formula at a top node of a Φ_i is at a top node of Ψ.

4. If Φ is a Prawitz tree and Ψ is obtained from Φ by *crossing out* some formulae at top nodes in Φ and superscripting them with exactly one integer, then Ψ is also a Prawitz tree.[2]

5. A Prawitz tree must be finite in length. That is, every Prawitz tree may be constructed by a finite number of applications of 1–4.

Call any part of a Prawitz tree of this form:

$$\frac{A_1 \ldots A_n}{B}$$

[2] All crossed out formulae in a Prawitz tree must have exactly one superscript. The crossing out of a formula indicates it as a hypothesis rather than a belief.

an *inference step*. Say that B is the *conclusion* of the inference step and that the A_i are the *immediate-premises*.

As I have defined a Prawitz tree, there is only one occurrence of a formula in a Prawitz tree that is not an immediate-premise of any formula. Say that such a formula is at the *bottom node* of the tree.

Example:

$$\dfrac{\dfrac{A}{B} \quad \cancel{C}^1}{D}$$

is a Prawitz tree. The formulae at top nodes are A and C, and the formula C has been crossed out. D is at the bottom node.

1.1.3 Above, Below and sub-dependence

- Say that an occurrence of A is *above* an occurrence of B if there is a sequence of occurrences of formulae $A_1 \dots A_n$ such that $A_1 = A$, $A_n = B$ and A_i is the immediate-premise of an inference step the conclusion of which is A_{i+1}.

It is easy to see that that A is above B in this technical sense exactly when A occurs physically above B in the Prawitz tree and A and B are in a common branch.[3] In example 1.1.2: A is above B and D; and A is not above C.

- Say B is *below* A when A is above B.

- Finally say that B *sub-depends* on A (in a Prawitz tree) when A is at a top node of the tree and is either above B, or is the same occurrence as B (that is, every occurrence of B sub-depends on every top node in the same branch as B).

So in 1.1.2: D sub-depends on A and on C; B sub-depends on A but not on C; and A sub-depends on A. Also C sub-depends on C.

1.1.4 Inference rules and side-conditions

A Prawitz natural deduction system is identified by its inference rules. Prawitz deductions are Prawitz trees that are constructed only in accordance with the rules.

[3] Where the branches are identified by following the horizontal lines.

An *inference rule* is a schema for inference steps. Every instance of an has this form:

$$\frac{A_1 \ldots A_n \quad \Phi_1 \ldots \Phi_m}{B}\ R\ \text{[side-conditions]}$$

where the A_i are formulae and the Φ_i are Prawitz trees. The side conditions are extra conditions of the structure of the Prawitz tree as a whole, which must be met in order for the inference step to be an instance of R.

The A_i and Φ_i the *premises* of the rule; the A_i are (also) the *singular premises* of the rule; B is the *conclusion* of the rule; and R is the *name* of the rule.

For example one of the rules for conjunction is this:

$$\frac{A \quad B}{A \wedge B}\ \wedge I$$

where '$\wedge I$' is the name of the rule. There are no side conditions on the rule AI. Another rule is this:

$$\frac{\begin{array}{c}\cancel{A}\\ \vdots \\ B\end{array}}{A \to B}\ {\to}I$$

In this rule the premise

$$\begin{array}{c}\cancel{A}\\ \vdots \\ B\end{array}$$

denotes any Prawitz tree Φ, the bottom node of which is B, and where there may be some occurrences of A at top nodes of Φ that are crossed out.

Inference rules of this sort *permit* instances of A to be crossed out. An inference rule like $\wedge I$ does not permit anything to be crossed out.

In chapter 9 I will use this rule:

$$\frac{\begin{array}{c}\cancel{A}\ \ \cancel{B}\\ \vdots \\ \bot\end{array}}{A|B}\ |I$$

the premise of this rule is a Prawitz tree the conclusion of which is $\perp$ and where occurrences of A or B at a top nodes are permitted to be crossed out.

What constitutes a legitimate side-condition?

In some of the inference rules I present later in the deduction the side conditions are quite long. Notice that the side conditions I present simply specify a decidable property of the premises of the rule. For example, here is a legitimate side condition for a hypothetical rule for a connective Con:

$$\frac{A}{Con(A,B)}\ ConI \quad \text{provided that } A \text{ does not depend on } B^4$$

it is a decidable matter, purely about the structure of the Prawitz tree, whether this side condition is met. Furthermore, the only way we have of expressing that inference rule for Con is by utilising a side-condition. All the side conditions I use in this thesis, no matter how complex, describe decidable syntactic properties of a Prawitz tree.[5] I claim therefore that there is nothing illegitimate about my side-conditions.

An example of a side-condition I think is illegitimate is this:

$$\frac{B}{A}\ R \quad \text{provided } A \text{ follows from } B \text{ in deduction system } X$$

where X is some deduction system. Firstly it might not be decidable whether the side condition is met. Secondly, in general, the side-condition of an application of R does not express a property of the tree containing that application. That is, we would have to look outside the Prawtiz tree to see if the side condition is met (i.e. to see if the Prawitz tree is a Prawitz deduction, see 1.1.8).

1.1.5 Rule applications

In a Prawitz tree, some inference steps may be labelled. Call an inference step a *rule application* when:

1. it is labelled by the name R of some inference rule.
2. it has a conclusion of the same form as the conclusion of inference rule that R names.
3. it may, or may not, also be labelled by some integers that are superscripts only of crossed-out formulae of the type permitted to be crossed out by the rule named R.

[4] See 1.1.6 for a definition of dependence.

[5] Except one, see the side condition for anaphora on page 148.

For example, this:

$$\frac{B}{A \to B} \to I(1)$$

is a rule application only if '1' superscripts only crossed out occurrences of A above B.

Say a rule application R occurs *below* an occurrence of A when A occurs above an immediate-premise of R.

Say a rule application R occurs *above* an occurrence of A when the immediate-premises of R occur above A.

1.1.6 Dependence

Dependence is a very important relation that can be defined only for Prawitz trees where every inference step is a rule application.

An occurrence of a formula B *depends* on an occurrence of A when

1. B sub-depends on A, and

2. either

   (a) A is not crossed out

   (b) A is crossed out and is superscripted by n, and there are no rule applications above B that both permit A to be crossed out and are labelled by n.

The supercripts and labels of rules applications indicate where a formula is discharged. More colloquially the definition above says that A depends on B when the labels do not indicate that B has been discharged by the time the deduction has reached A.

We can say generally that an occurrence of A *depends on* B, when that occurrence of A depends on an occurrence of B.

An occurrence of a formula B depends on a particular rule application when that rule application occurs above B.

Dependence here is not significantly different from dependence on the traditional interpretation of Prawitz trees. For both interpretations: an occurrence of A depends on B if B occurs undischarged at a top node in the deduction tree when the deduction has progressed as far as A.

Example

Here is a Prawitz tree using the two rules given above ($\wedge I$ and $\rightarrow I$):

$$\dfrac{\dfrac{\dfrac{\dfrac{\cancel{A}^1 \quad A^2}{A \wedge A}\,\wedge I}{A \rightarrow (A \wedge A)}\,\rightarrow I(1) \qquad A^2}{(A \rightarrow (A \wedge A)) \wedge A}\,\wedge I}{A \rightarrow [(A \rightarrow (A \wedge A)) \wedge A]}\,\rightarrow I(2)$$

Here are some facts about dependence in this deduction:

- The occurrence of $(A \rightarrow (A \wedge A)) \wedge A$ (in the second line from the bottom) depends on A, in particular it depends on the occurrences of A superscripted by 2.

- $A \rightarrow (A \wedge A)$ depends on only on the occurrence of A superscripted by 2.

- $(A \wedge A)$ depends on the occurrences of A superscripted by 1 and by 2.

- The bottom node of the tree depends on nothing.

Since all the side conditions on each inference rule are met (there are none to meet anyway), this Prawitz tree is a Prawitz deduction.

1.1.7 Dominance

The relation of *dominance* between occurrences of formulae is very important to natural deduction systems. I define it now specifically for Prawitz trees. One of the advantages of the Prawitz system is that the relation of dominance is quite simple.

- The immediate premises of an inference step dominate its conclusion.

- If Φ is a Prawitz tree, and the formula at the bottom node of Φ is the immediate premise of an inference step, then Φ dominates the conclusion of the inference step.

Effectively, any formula below a horizontal line in a Prawitz tree is dominated only by the formulae and deductions that occur immediately above the horizontal line.

The horizontal line is ambiguous

The horizontal line of the inference rules do not mean quite the same as the horizontal line of the Prawitz tree. The horizontal line of the inference rule indicates a relation of dominance, whereas the horizontal line of the Prawitz trees indicates a relation of succession.

The way I have defined things, the premise of an inference rule may be a Prawitz tree, but the immediate premise of a rule application is a formula.

In a Prawitz natural deduction, the conclusion of an rule application is dominated only by its immediate-premises and deductions the conclusions of which are its immediate-premises. In a sense this means that the history of the deduction is forgotten at each step as the conclusion of each rule application is derived only from its immediate predecessors (i.e. its immediate-premises). This gives the system much power and simplicity from the meta-logical point of view, and aids us greatly in defining and proving the necessary formal results.

1.1.8 Prawitz deductions and deduction-trees

A Prawitz *deduction* is a Prawitz tree Ψ such that

1. Every inference step is a rule application
2. The conclusion of every rule application is dominated by formulae and deductions of the same form as the premises of the inference rule by which it is labelled. Call such formulae or deduction the *true-premises* of the rule application.
3. The side conditions for each inference rule R are met at all the inference steps labelled by (the name of) R.
4. For every crossed-out formula A (or more precisely $\cancel{A}^n$) there is a rule application that occurs below A that both permits A to be crossed out and is labelled (in addition to a label naming a rule) by n.
5. No two rule applications are additionally labelled by the same integer.[6]
6. All rule applications are additionally labelled by at most one integer.[7]

[6]We could do without this condition, but to do so would lead to unnecessary complications.

[7]This condition is for simplicity and is not necessary, I add it here to neaten up the deductions. If I were writing a more general thesis on Prawitz deductions I would not include this clause.

The *conclusion* of the tree is the formula at the bottom node. The *premises* or *assumptions* of a Prawitz tree are formulae at top nodes of the tree that are not discharged. A *hypothesis* is a formula at a top node of a tree that is crossed out (and superscripted).

If C is at the conclusion of a Prawitz tree then any formula on which C depends is an assumption, any crossed out formula on which C depends is a hypothesis (crossed out formulae on which C does not depend are discharged hypotheses).

Some important simplifications of the language

1. From the way a Prawitz deduction has been defined, a true-premise A of a rule application is one of its immediate-premises, and a true-premise Φ of a rule application is such that one of its immediate premises is the bottom node of a deduction of the form/type of Φ.

   In other words, every rule application has the same form as the inference rule by which it is labelled.

   For this reason I shall use 'rule application' and 'inference rule' interchangeably (unless confusion may arise). Also I shall use 'premise' and 'immediate premise' interchangeably. I shall do this only in the context of Prawitz natural deductions (outlined above).

2. If C is the conclusion of a Prawitz tree and C depends on A, then call that Prawitz tree an *inference* from A to C. If C does not depend on any formula then that Prawitz tree may simply be called an *inference* of C.

3. If an inference from A to B that is also a deduction may be called a deduction from A to B.

The distinction between an *inference* and a *deduction* becomes important in later proof theoretic work where deductions are manipulated by rearranging their parts. The parts of a deduction are Prawitz trees and often they are deductions themselves, but not always (in the presence of assumptive rules such as the restart rule they may not be deductions). It is convenient to use a word similar to 'deduction' to describe these parts, I choose the word 'inference'.

1.1.9 Legitimate and illegitimate rule applications

Say a rule application is *illegitimate* when it occurs in a Prawitz tree if its conclusion is not dominated by premises of the same form as the premises of the inference rule by which it is labelled. A rule application is *legitimate* otherwise.

1.1.10 Appending Prawitz trees

I will refer to the *appending* (or adding) of one Prawitz tree to the end of another. Suppose Φ and Ψ are Prawitz trees, and suppose the conclusion of Φ is A. Then we may *append* Ψ *to the end of* Φ to obtain Ψ' by the following procedure:

1. If every occurrence of A at a top node of Ψ is crossed out (or there are no such occurrences at all) then $\Psi' = \Psi$. Otherwise,

2. Obtain Φ' by replacing uniformly all the integers occurring in Φ (as superscripts and additional labels) with integers that do not occur in Ψ.

3. Replace every occurrence of A at a top node in Ψ that is not crossed out with a copy of Φ'.

Usually Ψ' is a deduction if Ψ and Φ are, but not always (see section 11.1.3).

1.1.11 Discharge and empty discharge

Say that any rule application that permits the crossing out of A *discharges* A. We may talk also of particular occurrences of A being discharged by a rule application. Which occurrence of A is discharged depends on the labelling. For example, if an application of $\rightarrow I$ is labelled additionally by '1' then that application discharges all occurrences of A that are superscripted by '1'. Notice that there is no requirement that a rule application be labelled by any number, in such a case the rule application does not discharge any occurrence of A in the Prawitz tree. We may, however, still talk of A being discharged by such a rule application, more specifically we say that A is *empty-discharged*.

For rules such as these:

$$\dfrac{A \vee B \quad \begin{array}{c}\cancel{A}\\ \vdots \\ C\end{array} \quad \begin{array}{c}\cancel{B}\\ \vdots \\ C\end{array}}{C}\ \vee E$$

$$\cfrac{\begin{array}{c}\cancel{A}\\ \vdots\\ B\end{array}}{A \to B}\ {\to}I$$

at least one premise is a certain form of Prawitz tree rather than a single Prawitz formula.

For example a Prawitz deduction may contain rule application such as this:

$$\cfrac{\begin{array}{c}\vdots\\ B\end{array}}{A \to B}\ {\to}I$$

since the rule application is not labelled it does not matter whether or not there are any occurrences of A at top nodes above B.

Also worth noting is that this is a deduction:

$$\cfrac{\cancel{A}^{1}}{A \to A}\ {\to}I(1)$$

to see that this is a deduction note that A (occurring at a top node) depends on itself and the rule application of $\to I$ below it is labelled by its superscript.

Empty discharge becomes less mysterious, if we add this structural rule:

$$\cfrac{A_1 \ldots A_n}{A_i}\ S$$

and now we can make this deduction:

$$\cfrac{\cfrac{\cancel{A}^{n} \quad \begin{array}{c}\vdots\\ B\end{array}}{B}\ S}{A \to B}\ {\to}I(n)$$

and the premise of the rule application $\to I$ more obviously has the appropriate structure. I shall not use any structural rules like S because they are not necessary and add only tedious length to deductions and the formal results I shall obtain for Prawitz deduction systems.

1.1.12 Prawitz logical consequence

When using a Prawitz natural deduction system we may write $\Gamma \vdash A$ to mean that there is a Prawitz deduction the premises of which are among Γ and the conclusion of which is A.

So for example this

$$\frac{B}{A \to B} \to I$$

is a deduction witnessing that $B \vdash A \to B$ and that $\Gamma, B \vdash A \to B$ for any Γ. And this:

$$\frac{\cancel{A}}{A \to A} \to I$$

is a deduction witnessing that $\Gamma \vdash A \to A$ for any Γ (even empty Γ).

If Γ is empty then we may write $\vdash A$ instead of $\Gamma \vdash A$.

Finally, if Φ is a Prawitz tree that is a deduction that witnesses that $\Gamma \vdash A$, I shall call Φ a *deduction that* $\Gamma \vdash A$. That is, I shall write 'deduction that…' as short for 'deduction witnessing that…'.

1.1.13 Further remarks

Traditionally Prawitz deductions are not defined as I have done. I have given rules for generating a Prawitz tree and specified how the tree must be labelled and match the inference rules to be a deduction. Usually no distinction is made between Prawitz trees and Prawitz deductions, and the inference rules are given as rules for generating Prawitz deductions.

I define deductions in the more complicated way partly to make it easier to relate it to the deduction system of 1.2, and also so that the side conditions on the *restart rule* (chapter 8) make sense.

1.2 More natural deduction

1.2.1 Unnatural deduction

The branching structure of the Prawitz system can be a disadvantage if we wish to assert that it is matches the structure of actual deductions people make. To see why consider this deduction:

$$\dfrac{\dfrac{\dfrac{(A \wedge B) \wedge C}{A \wedge B} \wedge E}{A} \wedge E \qquad \dfrac{(A \wedge B) \wedge C}{C} \wedge E}{A \wedge C} \wedge I$$

Notice that $A \wedge (B \wedge C)$ is used twice, once to deduce A and once to deduce $B \wedge C$. This is not itself a problem, the problem is that in order to be used twice, it appears twice in the deduction. It would be more natural for the deduction to progress like this:

$$\begin{array}{cl} (A \wedge B) \wedge C & \\ \hline A \wedge B & \wedge E \\ \hline A & \wedge E \\ \hline C & \wedge E \\ \hline A \wedge C & \wedge I \end{array}$$

where the true premise of the lowest application of $\wedge E$ is $(A \wedge B) \wedge C$.

1.2.2 Dominance again

This difficulty is easily resolved by using a less restrictive notion of dominance.

First a definition:

> if this is an inference step of a Prawitz tree
>
> $$\frac{A_1 \quad \dots \quad A_n}{C}$$
>
> then (the occurrence of) A_1 and every (occurrence of a) formula above it is *to the left of* A_j and every formula above A_j for $i < j$.

Now we may define a more subtle dominance relation.

- An occurrence of A dominates an occurrence of B when
  1. The occurrence of A is either above the occurrence of B, or it is to the left of the occurrence of B.
  2. There is an occurrence of some C, below both A and B, that depends on everything that A depends on.[8]
- If Φ is a Prawitz tree, and the formula at the bottom node of Φ is the immediate premise of an inference step, then Φ dominates the conclusion of the inference step.

Dominance is extended only with respect to occurrences of formulae dominating other occurrences.

[8] I say 'everything' rather than 'every formula' as later I shall extend the notion of dependence so that occurrences of formulae may depend on rule applications as well as other formulae (see page 104 for a discussion of assumptive rule-applications).

With this more general definition of dominance the deductions can become more elegant. For example here is a deduction witnessing that $\vdash [A \to (B \to C)] \to [(A \to B) \to (A \to C)]$:

$$\dfrac{\dfrac{\dfrac{\dfrac{A \to (B \to C) \quad \cancel{A}^1}{B \to C}\, {\to}E \qquad \dfrac{\cancel{A \to B}^2}{B}\, {\to}E}{C}\, {\to}E}{A \to C}\, {\to}I(1)}{[A \to (B \to C)] \to [(A \to B) \to (A \to C)]}\, {\to}I(2)$$

The occurrence of B is dominated both by the occurrence of $A \to B$ above it and the occurrence of A in the other branch. Therefore we may apply $\to E$ to $A \to B$ without bothering to rewrite the A. To see that A dominates B note that C depends on A (A itself is everything that A depends on) and is below both A and B. Other examples are these deductions witnessing that $B \wedge A \vdash A \wedge B$:

$$\dfrac{\dfrac{B \wedge A}{A}\, \wedge E \qquad \dfrac{}{B}\, \wedge E}{A \wedge B}\, \wedge I \qquad\qquad \dfrac{\dfrac{\dfrac{B \wedge A}{A}\, \wedge E}{B}\, \wedge E}{A \wedge B}\, \wedge I$$

where B is dominated by $B \wedge A$.

The new dominance rule allows us to recall things into branch Φ which we have already deduced in branch a Φ' to its left. Think of a Prawitz deduction being generated from the top left, the deduction then proceeds downwards. Sometimes we begin a new subdeduction on the right which eventually meets below with the main stem (which began to its left). The dominance rule allows us to recall formulae we have deduced in a main stem into any subdeduction on the right. The condition on dominance itself is quite natural, although it can be tricky to formulate.

1.2.3 Why I stick with Prawitz

Although I think that the deduction system with the more general dominance condition is a better model of actual reasoning, I shall use Prawitz natural deduction (with the strict dominance condition of 1.1.7). I do this for three reasons

1. The difference between the two systems is not that great. With the more general dominance condition we can recall formulae deduced elsewhere. But we can do similarly in the Prawitz system, instead of recalling a past deduction we must re-deduce it. The difference between

the two is purely aesthetic as far as this thesis is concerned.[9]

2. The formal results I require are easier to formulate and prove in a Prawitz natural deduction system

3. Much of the literature to which the formal aspects of this thesis relates (e.g. Dummett's work) uses a Prawitz natural deduction system.

1.3 Other deductive concepts

1.3.1 Introduction and elimination rules

Let Con be some arbitrary connective.

An *introduction rule* for Con is an inference rule any application of which must contain Con in the conclusion, but need not contain any connective in any of the premises.

An *elimination rule* for Con is an inference rule any application of which has least one singular premise that contains Con, but need not contain Con in the conclusion.

For example, let $\vee$ be a binary connective and $\neg$ be a unary connective, then if this is a rule:

$$\dfrac{\begin{array}{c}\cancel{\neg A}\\ \vdots\\ A\end{array}}{A \vee B}$$

then it is an introduction rule for $\vee$, it is not an elimination rule for $\neg$. A rule of the form

$$\frac{A \quad {\sim}A}{B}$$

is an elimination rule for $\sim$. A rule of the form

$$\frac{C}{{\sim}A \vee C}$$

is an introduction rule for $\vee$ and for $\neg$. Finally, a rule of this form

$$\dfrac{\begin{array}{c}\cancel{{\sim}A}\\ \vdots\\ A\end{array}}{A}$$

[9] There are logical systems, e.g. that count the number of times a formula is 'used' in the deduction, for which the different dominance conditions make a significant difference. We shall not consider any such systems in this thesis.

is neither an introduction nor an elimination rule for anything.

There is a convention to label an introduction rule for Con as $Con\ I$, and to label an elimination rule $Con\ E$.

1.3.2 Major and minor premises

An elimination rule for Con has this form:

$$\frac{A_1 \ldots A_n \quad B_1 \ldots B_n \quad \Phi_1 \ldots \Phi_m}{C}$$

where the A_i are singular premises that contain Con and the B_i and Φ_i are other premises that need not contain Con.

The A_i are the *major premises* and the Φ_i and B_i are the *minor premises.*

For example in this famous elimination rule for disjunction:

$$\dfrac{A \vee B \quad \begin{matrix}\cancel{A}\\ \vdots\\ C\end{matrix} \quad \begin{matrix}\cancel{B}\\ \vdots\\ C\end{matrix}}{C}\ \vee E$$

the major premise is $A \vee B$ and the minor premises are $\begin{matrix}\cancel{A}\\ \vdots\\ C\end{matrix}$ and $\begin{matrix}\cancel{B}\\ \vdots\\ C\end{matrix}$. Also, in this elimination rule for $\rightarrow$:

$$\frac{A \quad A \rightarrow B}{B}\ \rightarrow E$$

the major premise is $A \rightarrow B$ and the minor premise is A. In the elimination rules for conjunction, e.g.

$$\frac{A \wedge B}{A}\ \wedge E$$

there are no minor premises and $A \wedge B$ is the major premise.

1.3.3 Maximal formula

In a Prawitz tree, a *maximal formula* is a formula that occurs both as the conclusion of an introduction rule and as the major premise of an elimination rule. For example if a part of a Prawitz tree looks like this:

$$\dfrac{\dfrac{A \quad B}{A \wedge B}\ \wedge I}{A}\ \wedge E$$

then that occurrence of $A \wedge B$ is a maximal.

1.3.4 Inconsistency

I shall define inconsistency only for logics the languages of which contain the special atomic formula $\perp$. A collection of formulae Γ is *inconsistent* iff $\Gamma \vdash \perp$.

In some cases it is possible to define inconsistency without reference to $\perp$ or any connective, but since all the systems I shall examine make use of $\perp$ I shall stick with the simple definition of inconsistency in terms of $\perp$.

1.4 Some formulaic concepts

1.4.1 Subformula

If A is atomic then A is a subformula of itself.

If Con is an n-ary connective then the subformulae of $Con(A_1 \dots A_n)$ are $Con(A_1 \dots A_n)$ and the subformulae of each A_i.

More formally:

$$Sub(A) = \{A\} \quad A \text{ atomic}$$

$$Sub(Con(A_1 \dots A_n)) = \{Con(A_1 \dots A_n)\} \cup Sub(A_1) \dots Sub(A_n)$$

1.4.2 Degree

The *degree* of a formula A is the number of occurrences of any propositional connective or quantifier in A.

For example if A is atomic then its degree is 0, but $A \rightarrow (A \rightarrow (A \wedge A)$ has degree 3.

1.5 Notation and quotation

It is convenient to use the letters $A, B, C \dots$ as variables over things that *express propositions* (i.e. thoughts and sentences, or more generally: utterances) rather than propositions themselves. I shall not discuss the nature and ontology of propositions.

I will use $\|A\|$ to abbreviate 'the proposition expressed by A'.

I shall occasionally use these little dots '. . . ' within quotation marks, for example:

> 'if. . . then. . . '

clearly I do not intend to quote the dots, I use them merely as a way of abbreviating this

> the binary sentential connective 'if then' taking its arguments immediately before and after the part 'then'.

I will use square quotation marks when using syntactic variables ($A, B \ldots$) within a quotation. So I write

> $\ulcorner$if A then $B\urcorner$

to abbreviate

> the utterance obtained by concatenating A and B appropriately with 'if. . . then. . . '

When presenting formal deductions, or writing sentences of a purely logical language, I shall omit the square quotation marks.

Chapter 2

Introduction

2.1 Historical analyticity

2.1.1 Leibniz' legacy

Kant's characterisation of an analytic truth was initially given in terms of a relation between concepts. An analytic judgement, for Kant, is a judgement where the concept of the predicate is contained in the concept of the subject. This idea is present also in Leibniz. If we can identify the predicate in a part of the subject then a judgement is analytic. An example of Leibniz' is that it is analytic that the part is less than the whole. The argument runs thus (note that when Leibniz writes 'part' he means 'proper part'):

> The part is equal to the part of the whole (for by an identity axiom, anything whatsoever is equal to itself). But that which is equal to a part of the whole, is less than the whole (by the definition of less). Therefore the part is less than the whole.[1]

This quotation presents an early formulation of the idea that the analytic truths follow from basic logical laws by applications of definitions of the relevant terms. So, by applying definitions relating to the terms 'part' and whole to a basic law identifying a part to itself we obtain the general truth that all parts are less than the whole.

A similar idea may be found in an alternate characterisation of analyticity given by both philosophers. An analytic truth is one the denial of which entails a contradiction. Definitions of this form are inherently circular as

[1] See page 340 of [?], the quotation is translated from Couturat ed. *Opuscules et fragments inédits de Leibniz*, 1903.

the only deductions of contradictions appropriate to this characterisation are deductions that do not depend on any synthetic truths (i.e. depend only on analytic truths). Of course, we can remove this circularity by defining a deduction that is independent of synthetic truths recursively, but this is tantamount to characterising the analytic truths in terms of deductions from basic analytic truths (which go without explanation).

The lack of a rigorous formal logic prevented Leibniz and Kant from formulating their view in a general and clear way. It was only until Frege's concept-script and the development of first order logic thereafter was the view more clearly expressible. The account of analyticity we derive from Frege and Carnap is essentially the same as that of Leibniz and Kant, especially in that it treats analyticity mostly in terms of a relation between concepts. The phrase 'the predicate is contained in the subject' is replaced, in Carnap's work, by talk meaning postulates and logic. The grounding of analyticity in concepts is brought to an extreme by Carnap in that he regards analyticity as being relative to a language, languages with different meaning postulates have different analytic truths. For Carnap, analyticity is a property of sentences rather than propositions.

In modern thinking more is made of a sharp distinction between sentences and propositions. Only propositions bear truth values and since we talk so frequently of 'analytic truths', it is natural to regard analyticity as a property of propositions rather than sentences. This poses a problem if we wish to maintain the traditional Leibniz-Kant view of analyticity as arising from a relation between concepts, in modern terms the Leibniz-Kant view is that analyticity arises from a relation between *meanings* (hence the catchphrase 'an analytic truth is truth virtue of its meaning'). The problem lies in the fact that propositions are independent of meanings and concepts, no matter how we structure our concepts or choose our meanings, the situations where a proposition is true remains unchanged. A modern problem for a theory of analyticity is to provide a method of reconciling the intuition that an analytic sentence is 'true by meaning' with facts relating to the independence of meaning and propositions. Boghossian's attempt (in [**?**] among other works) at this is to treat analyticity as an epistemological property, analytic propositions are ones which can be *known* to be true only by knowing the meanings of the terms. In section 4.2 I object to Boghossian's position, in particular I object that analyticity is not an epistemological property.

2.1.2 Wittgenstein

This brief discussion of the history of analyticity would not be complete without at least mentioning the view provided by Wittgenstein in the *Tractatus*. The approach to characterising analytic truths followed by Kant, Leibniz, Carnap and many modern philosophers has a strong syntactic flavour. That is, the analytic truths are those that are *deducible* from basic analytic truths (e.g. logic) and definitions of meaning (e.g. Carnap's meaning postulates). Wittgenstein's characterisation is semantic, in 4.46 he defines a tautology as a proposition that 'is true for all the truth-possibilities of the elementary propositions'.[2] Interestingly, Wittgenstein was the first, as far as I am aware, to provide some explanation (other than an appeal to the light of reason) of why the basic axioms of logic are analytic (at least, he can account for why they are tautologies):

> The sign which arises from the co-ordination of that mark "T" with the truth-possibilities is a propositional sign. (4.44)

So a symbol is defined by its truth table and the tautologies are those propositions the truth functional structure of which is such that they are true on all combinations of truth values for the atomic propositions. From this Wittgenstein concludes that a tautology is senseless in that it provides us with no information.

Wittgenstein's account is problematic mainly because truth tables are insufficient for handling the vast body of logical truths expressible only in first (or higher) order languages. A more advanced account that follows the semantic line is given by Peacocke in [**?**]. In addition of specific objections to Peacocke's position, I have a general worry about approaches that, like Wittgenstein's, characterise analytic truths in terms of a semantics: I worry that assuming a semantics begs the question of which sentences are analytic. For example the only appropriate interpretation for disjunction and conjunction on a truth-table semantics yields the law of excluded middle as a tautology. Indeed, the law of excluded middle, if genuinely a logical law, is a prime candidate for being analytic. So it seems that in assuming a semantics in order to explain what the analytic truths are, we come dangerously close to assuming, in advance of the explanation, what the analytic truths are.

---

[2] I take my quotations from Ogden's translation [**?**], a hypertext of which can be found at www.kfs.org/ jonathan/witt/tlph.html

2.2 Criticisms of Analyticity

I wish to provide at least part of a defence of the truth-by-definition account of analyticity in the tradition of Leibniz, Kant and Carnap. There are two main lines of criticism to such an account (I consider them all in detail in chapter 4).

1. Firstly, there are objections to what the tradition takes analyticity to be. Philosophers (e.g. Harman and Boghossian) have taken the true-by-definition approach to be highly objectionable, the basic objection being that propositions are not true or false depending on meaning (nor definition), they are true or false depending on the way the world is. I outline these objections in more detail in section 4.2. Such objections are easily answered by noting that the Leibniz-Kant-Carnap tradition takes analyticity to be strictly speaking a property of sentences that express propositions and not the propositions themselves. It is not propositions, or truths, that are analytic but the sentences that express them. If a sentence can be deduced from the rules that define (some of) its terms then it is analytic, and furthermore since the rules are definitional, it will express a true proposition. The exact details of how this is to work are the subject of the second sort of objection.

2. Secondly there are more serious objections about the possibility of giving a coherent account of definitions. The two main questions posed by such objects ask for an account of how and when these definitions are made, and what exactly counts as a definition.

    (a) Quine's famous objection (in [**?**]) to the analytic-synthetic distinction applies mainly to a theory of analyticity that follows the lines of Carnap and Frege. Quine objects that there is no way to tell what the definitions are (the reason there is no way to tell, Quine concludes, is that there never were any such definitions made).

    Quine charges the theory of Carnap with a circularity similar to that I outline above against the alternative Leibniz-Kant characterisation of analyticity (in terms of a denial entailing a contradiction, see page 34). Quine argues that the question of which truths are analytic includes the question of which sentences really are the meaning postulates. For example, if we try to identify the meaning postulates in terms of synonymy then we had better explain how these synonymies arise otherwise we may as well leave the notion of analyticity unexplained. Effectively, Quine challenges Carnap's account of analyticity to state what the facts are that determine which meaning postulates are appropriate to a certain language. Quine argues that this cannot be done without relying on a notion, such as synonymy, that itself cannot be explained without giving a theory of analyticity. Quine concludes that there is no fact of the matter as to which meaning postulates are correct, and ultimately that there is no fact of the matter as to which sentences are analytic and so that the whole analytic-synthetic distinction distinguishes nothing.[3]

Although Quine does not argue specifically against the idea that at least logical truths are analytic we can easily extend his arguments. We can question when the axioms of, say, first order predicate logic with identity were set as definitions of the meanings of the logical terms. If no such definitions were made there is no reason to suppose them as analytic.

(b) A more serious objection questions the reliability of making definitions. At least, the reliability definitions that do the work needed to derive the consequence relation of a basic logic.

We can, the objection goes, reliably define new terms by introducing one new term to mean another. For example we might introduce a new word splurg as follows:

(†) x is splurg iff x is unmarried and x is male

And now 'splurg' means the same as 'bachelor' by definition. But to carry out such a definition we need a logic to understand the terms 'iff' and 'and' and perhaps even 'x'. But it is exactly such a logic that the truth-by-definition account of analyticity is attempting to explain. So a non-circular account of analyticity cannot base itself on definitions like †. But, the objection continues, there is no other way of making definitions that is reliable. Other apparent ways of defining terms (e.g. by writing axioms for them) unreliably allow pretty much any sentence to be definable as true, not just any axiom is a legitimate definition. For example we do not define 'bachelor' further by laying down (in addition to them being unmarried men) an axiom that all people are bachelors, certainly this would neither make 'all people are bachelors' analytic nor true.

[3]By stronger arguments Quine reaches his thesis that meaning itself is indeterminate

2.3 The beginnings of an answer

The theory of implicit definitions is a more modern attempt to answer this objection. The idea is that logical terms may be defined implicitly by their inference rules. Since we draw inferences all the time the theory of implicit definition hopes to answer both the question of what a definition is and the question of how and when the definitions are made.

The implicit definition theory requires us to identify the basic logical laws not in terms of axioms but in terms of inference rules. For example, we commonly reason with these inference rules:

$$\frac{A \quad B}{A \wedge B}\ \wedge I \qquad \frac{A \wedge B}{A}\ \wedge E \quad \frac{A \wedge B}{B}\ \wedge E$$

and somewhere along the line these rules define the connective $\wedge$. Anything deducible from rules that implicitly define terms may then be said to be analytic. For example, we can use the rules that implicitly define conjunction and negation to deduce that not-(A and not-A.

For such a theory to work we must respond to the objections: how are the implicit definitions made?[4] And how does the account rule out spurious definitions?

For one thing it is to be explained why the inference rules above, for conjunction, is definitional but this inference rule is not:

$$\frac{x \text{ is a mammal}}{x \text{ has a liver}}$$

After all, the inference that something is mammal from it having a liver is truth preserving and perhaps made without much thought by people with some knowledge of biology. To respond to Quinean objections, the implicit definition theory much be able to give an account of the difference between an definitional inference rule and an inference that happens to be truth preserving.[5]

More seriously, defining a term by inferential practices seems highly unreliable, not just any inferential practice will make for a satisfactory definition. A famous illustration of this is Prior's connective tonk which has these

[4] I shall argue that one way of fleshing out the theory of implicit definitions, in terms of inferential roles and practices, fails to answer this question adequately.

[5] A natural line is to make some metaphysical or semantic distinction between a mere inference pattern and a *rule*. The inference patterns for conjunction are *rules* but the pattern for mammals and livers are not *rules*. But exactly what a 'rule' is that distinguishes it from any other inference pattern is hard to define. Indeed, arguments discussed in in [**?**] suggest that such a distinction is impossible.

inference rules

$$\frac{A}{A\ tonk\ B}\ tonk I \qquad \frac{A\ tonk\ B}{B}\ tonk E$$

The problem is that these rules allow anything to be deduced from anything. So if the tonk rules were definitional then anything could be deduced from anything *by definition*.

This problem of reliability applies to any theory of analyticity that relies on inference rules to define the meanings of logical terms.

2.4 The spirit of inferential roles, my goal

Almost all attempts at providing an account of analyticity of which I am aware base themselves on the idea that a set of inference rules for a logical connective can define it in some way. To make such a theory workable we must give an account of how these definitions are made, and more importantly, what inference rules can count as definitions.

The objection to these theories that I will spend most energy answering is this:

> The only frameworks for making definitions reliably (so that no spurious definitions arise), are explicit definitions that merely define one term to mean another (usually a significantly more complex and descriptive one). Such definitions already require a logical language to work, and so are incapable of accounting for the analyticity of logic (without being circular). The theory of implicit definition solves the circularity problem, but it is not reliable, spurious definitions arise.

To respond to this objection we need to give an account of what inference rules make for legitimate definitions, when we present such an account we must give some reason to believe that it rules out any spurious definition.

Mainly I shall be concerned with the logical consequences of a famous proof theoretic account of what inference rules can make a genuine definition. I present and motivate this account in chapter 3. The account is that any inference rules that are in *harmony*, in Dummett's sense,[6] may be used as an implicit definition. At least for the main work of the thesis I shall not argue that only harmonious inference rules are useable as implicit definitions, but I shall suggest that inference rules that are not harmonious are suspect.

[6] The same idea can be found in writings of Gentzen and Hacking. Readers impatient for a definition of harmony should refer to the entry in the index.

The large logical portion of the thesis shows how we can obtain rules for interesting logics (in particular classical logic, and free logics) all satisfying a harmony constraint.

I shall not spend much space answering the question of when and how implicit definitions are made. In chapter 14 I shall postulate a theory of how an inferential definition could actually be made, but I shall not discuss it deeply. It is interesting to note that it follows from theory I give in chapter 14 that *only* harmonious inference rules can make for implicit definitions.

2.5 A theory of analyticity

Here is definition of analyticity that I hope this thesis goes some way to defending.

My aim is to revive the catchphrase 'analytic sentences are true by definition' or 'analytic sentences follow from definitions alone'. I interpret the catchphrase in this way:

> Analytic sentences, by definition, express true propositions (express propositions that satisfy the property of truth).

It is helpful to use a more general notion of analyticity so it can be applied to inferences:

> The premises and conclusion of an analytic inference, by definition, express propositions that satisfy a relation of validity.

I will now clarify how this could be the case.

Let s be a sentence and S be a collection of sentences. Assume also, that s and the members of S are interpreted on the conventional meanings of the terms.[7]

> s is *analytic* for S if there is (in abstract) a deduction such that
>
> 1. the premises are in S
> 2. the conclusion is s
> 3. every step of the deduction is in accordance with a definition

[7]So for example if s is 'grass is green or it is not' then the words of that sentence are interpreted as they conventionally are, rather than e.g. such that s expresses the proposition that snow is white.

An inference is analytic when its conclusion is analytic for its premises. A sentence s is analytic when an inference of s from an empty collection of premises (i.e. no premises) is analytic.

As we shall see, there are two sorts of definition. Some definitions work by presenting a sentence, e.g.

(†) Anything is a bachelor iff it is an unmarried man

other definitions work by presenting some rules of inference. For example

$$(\ddagger)\quad \frac{A \quad B}{A \text{ and } B} \qquad \frac{A \text{ and } B}{A} \qquad \frac{A \text{ and } B}{B}$$

So 'every step of the deduction is in accordance with a definition' comes down to a condition that

1. every premise of the deduction is in S or a sentential definition (like †)
2. every inference step of the deduction has the form of an inferential definition (like ‡).

But why should something that follows from definitions alone in this way be true? More generally, why should the proposition expressed by s follow validly from the propositions expressed by S just because s is analytic for S? The short answer to these questions is:

Because they follow by definition!

a curt response, but not entirely unreasonable when used in conjunction with an answer to the problems of the reliability of implicit (i.e. inferential) definitions.[8] The longer answer is given by my postulated theory of chapter 14.

2.5.1 Some facts about the analytic

To help provide an idea of what my view of analyticity is, here are some claims about analyticity. The claims of this section are not definitive and I am prepared to be convinced otherwise. I add this section because to help the reader understand what my broad motivations are and why I develop the particular logics of later chapters.

[8] A theory of analyticity which says what a definition is but cannot explain exactly why they work is better than one which cannot even say what the definitions are.

- Any instance of the law of excluded middle $A \vee \sim A$ (provided that A really does express a proposition) is analytic. It also is true, that is, it expresses a true proposition. The true proposition that $A \vee \sim A$ expresses is not true because of the definitions we make about $\vee$ and $\sim$.[9] However, the meanings of the terms $\vee$ and $\sim$ are such that $A \vee \sim A$ expresses a true proposition.

- 'Analytic' is a predicate of sentences, and so analyticity is relative to a language. However, roughly the same propositions are expressed by analytic sentences of various languages. In particular, the propositions expressed by logical truths are expressed by analytic sentences in any language. I claim this is because a speaker could not even develop an advanced language without defining a basic logical language ('and', 'or', 'not', 'if...then...' etc.). I postulate a theory to back up these claims in chapter 14.[10]

  It is perhaps even a truism that there are no analytic sentences of one language the translations of which into another language are not analytic. I suspect that it is a condition on something being a translation that it preserve analyticity. But I am unsure if this is correct.[11]

- The sentence 'If Hesperus exists then Hesperus is Hesperus' is analytic, for it follows from the logic of identity. However, since 'Hesperus is Phosphorus' does not follow deductively from the definitions or events that fixed the references of the two names, 'Hesperus is Phosphorus' is not analytic.

My theory of analyticity extends, in this thesis, only as far as logical truths. So I do not address directly the question of how a sentence such as 'all bachelors are unmarried' be analytic, it should however be clear what my account of such sentences is. 'All bachelors are unmarried' is analytic as it

[9] I cannot provide an explanation of why any particular proposition is true.

[10] My view is the contrary of Salmon's claims at the beginning of Appendix B of [?]. I hold that epistemic and deductive (logical) properties apply primarily to sentences (utterances), whereas semantic properties such as truth and validity apply primarily to propositions.

[11] My thought is this, suppose I say 'I am here' and consider how this is to be translated into the language English^{-I} that is just like English except it does not possess any first personal pronouns (like 'I'). It seems like a good translation of 'I am here' (were I to utter it) is 'Gabbay is here'. Perhaps the translation is not perfect and the two sentences do not express the same proposition, but if they do, then 'I am here' is analytic (I claim) but its translation into another language is not.

follows from logic and the explicit definition 'anything is a bachelor iff it is an unmarried man'.

Here are some further claims about analyticity which shall not figure in this thesis (except perhaps my belief that 'something exists' is analytic, see 11.3). They are, however, claims form part of a natural extension of my account of analyticity:

- Not all analytic truths are necessary. This claim does not affect my thesis as I concern myself to explain logical truths, and these are necessary. However I think that we could, on learning that there is a unique inventor of the zip, stipulate 'Julius invented the zip' and thereby define the name 'Julius'. 'Julius invented the zip' is analytic but not necessary.[12]

- 'I exists' and 'something exists' are analytic, as is 'I am here now'. The analyticity of these have been disputed, cunning counterexamples are occasionally passed around where someone may utter, seemingly falsely, 'I am here' with an intention to deceive. For example I may, pretending to be my brother, say 'I am here' to get someone to believe that my brother is here. The thought is that in such a case 'I' refers to my brother and not me, and so 'I am here' is false. I am inclined to deny that, the utterance of 'I am here' in such a case is still true, but its delivery is deceptive. Alternatively I could argue that the use of 'I', in the deception, is not on its conventional meaning.

These claims follow from my considerations in chapter 14 and do not follow from anything I present before then. I state these claims here because some of the logics I develop before chapter 14 are designed to be compatible with the considerations of that chapter. In particular they are designed to be compatible with the claims I make here.[13]

[12]Look at sections 8.1.1and 11.3 for my use of a 'weak' inference rules and their relation to the strict conditional, the stipulation that 'Julius' denote the inventor of the zip may be analysed like this:

$$\frac{\text{someone invented the zip}}{\text{Julius invented the zip}}\ \text{Julius-I}$$

where the rule Julius-I is weak. Intuitively a weak rule is a rule that implicitly depends on some contingent fact. In this case the contingency is whoever happened, in actuality, to have invented the zip (so as to be the object of reference of the rigid designator 'Julius').

[13]Note that I write 'be compatible' and not 'entail' or 'be entailed by'.

Chapter 3

When is a definition a definition?

3.1 The core of the implicit definition approach

My goal is to provide an adequate basis to make this claim: a logical connective has the semantics it does because it is defined so that it has it. For example I favour a theory of analyticity that states that e.g. the conjunction sign is defined so that it expresses the truth function that maps to truth only when both conjuncts are true.

This sort of account seems open to a charge of circularity: how can we define 'and' to mean what it does without already possessing a conjunction in our language, how can we have a definiendum without already possessing a definiens? The answer is that the model of explicit definition of this form (to use the equality symbol somewhat loosely):

definiendum = definiens[1]

is only one of at least two ways of making a definition. An alternative way is to make the definition implicitly by stating some rules or propositions involving the term to be defined and stipulating that this term express whatever makes those rules/propositions valid/true. For example we may present these rules

$$\frac{A \quad B}{A \wedge B} \qquad \frac{A \wedge B}{A} \qquad \frac{A \wedge B}{B}$$

and define thereby $\wedge$ to have the (simplest) meaning that makes them valid.

[1]For example: bachelor = unmarried man.

Such a definition is called an *implicit* definition and has a definiendum but no definiens (as opposed to an *explicit definition* which involves an definiens).[2]

This is the core of a truth-by-definition tradition of analyticity (a tradition I support) and, as argued above, in order to make it work we must provide an account of what exactly a counts as a definition. In particular, in the light of the example of Prior's connective tonk, we must give an account of which inference rules can be used as implicit definitions.

It is helpful to begin by discussing explicit definitions and then generalise to implicit definitions.

3.2 Explicit definitions and conservativeness

Conditions on an explicit definition being successful are not so easy to give. Consider this explicit definition of the new term 'splurg'

(†) Grass is blue and x is splurg iff x is a human male

the 'definition' does not define 'splurg' to mean 'human male', it is not a definition at all. † entails that if anything is a human male then grass is blue, a conditional such as that cannot be true by definition.

---

[2]There is an ambiguity in the phrases 'explicit definition' and 'implicit definition', the term 'explicit' could either be referring to a property of the definition or a property of the action of defining. Thus a definition might be called explicit if it an act of defining has been carried out explicitly.

So there are at least four things a definition could be:

1. An *explicit* definition made *explicitly* in our actions
2. An *explicit* definition made *implicitly* in our actions
3. An *implicit* definition made *explicitly* in our actions
4. An *implicit* definition made *implicitly* in our actions

For example, if I write down these rules:

$$\frac{A \quad B}{A \wedge B} \qquad \frac{A \wedge B}{A} \qquad \frac{A \wedge B}{B}$$

and say 'I hereby define $\wedge$ by these rules', then I have explicitly made an implicit definition. Alternatively if I simply use the rules without stating explicitly that that they are part of a definition then I have made (or am making) an implicitly made an implicit definition.[3]

To sum up, a definition is explicit or implicit depending on whether there is a definiens (if there is no definiens then the definition is implicit). But a definition may be explicit or implicit in another sense where the act of making the definition is explicit or implicit in our actions. As far as I can see, the literature on the subject, at least with respect to the discussion of inferential roles, distinguishes only between explicitly made explicit definitions and implicitly made implicit definitions.

If we were to regard † as a definition then we would be entitled to conclude, since I am a human male, that grass is blue. Thus † does not merely define a new term, it entails a false proposition. The problem with † is that we can use it to derive new information that has nothing to do with 'splurg' that we could not previously derive.

Such considerations have led some philosophers hold that the condition of legitimacy of a definition is *conservativeness*. One must be careful with use of the term 'conservative', it is a binary predicate relating a theory T to an extension of T. For the purposes of our discussion we will have to define conservativeness more generally as a relation between logical consequence relations, my thesis is about logical consequence relations rather than theories.

3.2.1 The conservative condition

Let $\vdash_1$ be a logical consequence relation and $\vdash_2$ be a logical consequence relation that includes $\vdash_1$. That is, if $\Gamma \vdash_1 A$ then $\Gamma \vdash_2 A$. Since every consequence in $\vdash_1$ is a consequence in $\vdash_2$ we may call $\vdash_2$ an *extension* of $\vdash_1$. Let L_1 be the language of $\vdash_1$ and L_2 be the language of $\vdash_2$.[4] Suppose further that the formulae in Γ_1 and the formula A_1 are in the language L_1.

Now, $\vdash_2$ is a *conservative extension of* $\vdash_1$ (or, $\vdash_2$ is *conservative over* $\vdash_1$) when, if $\Gamma_1 \vdash_2 A_1$ then $\Gamma_1 \vdash_1 A$. In other words, $\vdash_2$ is conservative over t_1 when any deduction in t_2, which has its premises and conclusion all in the language of $\vdash_1$, could have been carried out in $\vdash_1$.

Now we can present at least necessary conditions for the legitimacy of a definition. Let $\vdash_2$ be obtained from t_1 by the addition of a new term and its defining axioms or rules. Then then the defining axioms/rules are legitimate definitions exactly when $\vdash_2$ is conservative over $\vdash_1$.

3.2.2 A conservative working hypothesis

I postulate that conservativeness is both necessary and sufficient condition on the legitimacy of definitions because it fits with intuition about what definitions do, it is very successful as a formal condition (e.g. see the discussion of extensions by definition in [**?**]), and I know of no genuine alternative.

But this alone is of little philosophical value. What we are looking for is a condition on the definition itself, something about the way the definition

[4]For example, $\vdash_1$ may not be formulated in a language that contains 'splurg', whereas $\vdash_2$ is an extension of $\vdash_1$ so as to account for 'splurg'.

is made, that makes it legitimate. In particular, we wish to find a property a definition might have that explains why it is conservative.

3.2.3 The philosophically helpful version

Intuitively, if an explicit definition satisfies these criteria then it is legitimate:

(i) the definiendum is not contained in the definiens (i.e. the definition is non circular)

(ii) the definiendum is a single term of a certain syntactic category[5]

(iii) the term is totally new to the language.

We can now sketch a proof that this intuitive condition is enough to ensure conservativeness. I do it here for definitions of the form 'definiens iff definiendum'.[6] We may always replace the definiens with the definiendum without loss of validity, thus in any deduction $\Gamma \vdash_2 A$, where neither Γ nor A contain the definiendum, we may obtain a deduction that $\Gamma \vdash_1 A$ that makes no use of the definition and contains no occurrence of the definiendum. We do this by replacing any occurrence of the definiendum in the deduction by the definiens. Because of (iii) the only way the definiendum can get into the deduction is by use of the explicit definition, and so because of (i) our uniform replacement of the definiendum by the definiens will remove the definiendum completely from the deduction. Further, because of (iii) the result of this replacement is that any instance of 'definiendum iff definiens' becomes 'definiens iff definiens', which is trivially true.

(i), (ii) and (iii) are philosophically more useful conditions than the condition of conservativeness, because they can be given as a general property of good definitions that anyone can test for. For example, it is more plausible that we have an innate capacity for making definitions and checking that (i), (ii) and (iii) hold (thereby coming to know that they are legitimate definitions), than that we have an innate capacity for checking conservativeness of extensions. (i), (ii) and (iii) describe simple syntactic properties whereas conservativeness is a complex one.

[5]That is, we just present the term on the left and say whether it is a predicate or a propositional connective etc. and then say what it is to mean on the right (by presenting a complex from the same syntactic category).

[6]For definitions of the form 'definiens = definiendum' we must either ensure that there is an additional existence condition on the definiendum, or use a logic that can handle empty reference (I present such a system in section 11.3).

I do not conclude from this that satisfying (i), (ii) and (iii) is the only way by which something may be a legitimate explicit definition. But I suggest that anything that does satisfy (i), (ii) and (iii) is a legitimate definition, and that anything that does not is suspect (and we have reason to worry that it is not). The point of (i), (ii) and (iii) is that it provides us with a way of differentiating legitimate from potentially illegitimate definitions solely on the basis of the structure of the definition and not on its consequences. So for example the fault in the 'definition' † above is that 'human male' is not new, it already has a meaning.

If we could build a theory of analyticity based only on explicit definitions that satisfy (i), (ii) and (iii), then we have responded to the reliability worry (for we can be sure that the only definitions that enter our account of analyticity are legitimate ones). But of course for reasons that Quine has famously given explicit definition will not do to ground a theory of analyticty. Firstly, an explicit definition contains the term 'iff' or '=', and those are among the terms for which I seek to give an account. Thus an appeal to explicit definitions (where definiendum and definiens are linked by a logical connective) would be of little help. Secondly, in order to define the definiendum by means of an explicit definition we must already be able to express something with the same meaning: the definiens. So an explanation of our acquisition of a logical language in terms of explicit definitions would presuppose that we already acquired a logical language. This is the essence of Quine's argument in *Two dogmas of Empiricism*.

3.3 Legitimacy of implicit definitions and conservativeness

It is natural then to extend the conclusions of the discussion of explicit definition that of implicit definitions. If $\vdash_1$ is extended to t_2 by means of an implicit definition, then the extension must be conservative. But once again, this is unhelpful unless we can find something along the lines of (i), (ii) and (iii) for implicit definitions.

I think (ii) and (iii) may apply equally well to implicit definitions and that at least part of the analogue of (i) is that among the rules for the connective in question are introduction and elimination rules for it. But more is required. The explicit definitions we considered were all single equivalences or equlaities. So for example, something like this would not do as an explicit definition of $\ulcorner F \urcorner$ in terms $\ulcorner G_1 \urcorner$ and $\ulcorner G_2 \urcorner$

If x is G_1 then x is F

x is G_2, if x is F

unless it could be reformulated into something in the form of

x is G iff x is F

where G is obtained from G_1 and G_2 (the best way for this to happen is when G_1 and G_2 are both G).

Now the inference rules for a connective may be seen as implication clauses without the 'if... then...' where the introduction rules are left-right and the elimination rules are right-left (the new term beign on the right). We must find a suitable condition that shows we may also think of the rules collectively as a single 'if and only if'.

3.3.1 Harmony

A condition that appears to achieve what we want is *harmony* or normalisation. Normalisation is a proof theoretic property. A logic normalises (is harmonious) when any deduction in that logic can be reduced to a deduction, with the same premises and conclusion, in *normal form.* A deduction is in normal form when it involves no introduction and subsequent elimination of any logical connective. For example, consider this Prawitz tree:

$$\dfrac{\dfrac{\begin{matrix}\vdots & \vdots \\ A & B\end{matrix}}{A \wedge B}\wedge I}{A}\wedge E$$

this involves the introduction and subsequent elimination of a conjunction. In order to introduce the conjunction both A and B are deduced, the conjunction $A \wedge B$ is then used to conclude A. It should be easy to see that this entire deduction can be replaced by the initial deduction of A.

$$\begin{matrix}\vdots \\ A\end{matrix}$$

In this deduction the premises and the conclusion are the same, but the introduction and subsequent elimination of conjunction has been gone.

'Harmony' has at least two senses.

1. A connective may be said to be in harmony when where is a pleasing symmetry to its introduction and elimination rules (like those for conjunction).

2. Alternatively a connective may only be said to be in harmony in the context of logical consequence relation: a connective Con is in harmony relative to $\vdash$ (a logical consequence relation), when any deduction in $\vdash$ involving an introduction and subsequent elimination of Con in a deduction in $\vdash$ may be replaced by a deduction with the same premises and conclusion but which lacks that (unnecessary) introduction and subsequent elimination of Con.

If we add Con to a logical consequence relation $\vdash$ so that Con is harmonious in its second sense then Con makes for a conservative extension (defined below). For if Γ and A do not contain Con then any use of Con in a deduction that $\Gamma \vdash A$ must have been introduced by a rule for Con, and then it must be eliminated at a later point so that Con does not appear in A. But since Con is in harmony, this need not ever happen, and so Con need not appear in a deduction that any deduction that $\Gamma \vdash A$.

If every connective is in harmony in its second sense the logic as a whole normalises.

I will be concerned only with harmony in its second sense, and since I will expect every connective in the logics I develop to be in harmony I use the terms 'harmonious' and 'normalises' interchangeably. That is I will write of whole systems of connectives (i.e. logics) being in harmony, meaning that they satisfy normalisation theorems.

Intuitively, when a connective is in harmony (in its second sense), then the conclusions we can draw from the use of a connective are no more than what was required to introduce it. For example, the consequences of a conjunction are no more than what is required to introduce it: the rules for conjunction require that A and B be deduced in order to introduce $A \wedge B$, but from A and be we can deduce only that A and that B (back where we started). Certainly Prior's famous connective *tonk* is not in harmony in any logic that contains it, (see pages 80, and 93 for the rules and discussion of tonk).

To verify that the condition of harmony is what we are looking for we must check that it entails conservativeness, and I think it would be a bonus if a proof that harmony entails conservativeness uses follows a similar structure to a proof that (i-iii) entails conservativeness.

3.3.2 Conservativeness and harmony in the first sense

Now, it does not follow that if a connective is harmonious in the first sense that it will be a conservative addition to a logic.

For example, the logic with only one connective *tonk* and these rules:

$$\frac{A\ \mathit{tonk}\ B}{B} \qquad \frac{A}{A\ \mathit{tonk}\ B}$$

is consistent, it has no theorems, but if we extend it by adding 'if... then... ' with its familiar rules (modus ponens and conditional proof) the logic becomes inconsistent.[7]

Another example is how conjunction, when added to relevance logic is not conservative. Relevance logic distinguishes between this structure:

$$\begin{array}{c} \vdots \\ A \end{array}$$

and this structure

$$\begin{array}{cc} \vdots & \vdots \\ A & B \end{array}$$

The first is relevant to A but the second is not (as it contains an irrelevant inference of B). But with conjunction the two structures become inter-deducible as by adding

$$\cfrac{\cfrac{\begin{array}{cc} \vdots & \vdots \\ A & B \end{array}}{A \wedge B}}{A}$$

we can convert the second irrelevant (to A) structure into the first (which is relevant to A).

Perhaps the most infamous example of all is this:

> $\ulcorner x$ is in the set of all $F\urcorner$ expresses the local-validity/truth of: Fx

which, at face value, entails the existence of a certain kind of abstract object and, more notably, it entails an inconsistency when in the presence of a negation. The inconsistency is Russell's paradox, see 101 for a fuller discussion of this.

[7] The rules for 'if... then... ' allow a derivation of $\ulcorner$if A then $A\urcorner$ and using the rules for *tonk* from that there is a two step derivation of anything we like:

$$\cfrac{\cfrac{\text{if } A \text{ then } A}{(\text{if } A \text{ then } A)\ \mathit{tonk}\ B}}{B}$$

We could look for more conditions to (i-iii) so that the addition of † is guaranteed to be conservative. For example, that the exact structure of the deduction (as it may be written down) is not important to the logical consequence relation to be extended (to handle the case of relevance logic). We could resolve these problems by being very strict on the structural properties of the logical consequence relations over which we define our connectives, as Hacking is in [**?**]. If we go such a root then we must rule out, say, relevance logic, as a possible logic before we even begin analysing what logical connectives we use. This is too strict for me.

3.3.3 Conservativeness and harmony in the second sense

The solution, I think, is to look at how conditions (i-iii) entail conservativeness in the case of explicit definitions. To show that an explicit definition satisfying (i-iii) is conservative we argued that the definiendum may be replaced throughout by the definiens in a deduction that $\Gamma \vdash_2 A$. Suppose that the definiendum does appear in the deduction that $\Gamma \vdash_2 A$. Remember we are assuming that Γ and A do not contain any use of the definiendum. With this assumption in mind it follows that the only way the definiendum could appear in the deduction is if it were introduced by appeal its explicit definition (as the definiendum does not appear in Γ). Furthermore, since the explicit definition is, by assumption, the only rule that deals with its definiendum, the only way the definiendum could, subsequently, fail to appear in the conclusion A is if it was removed (i.e. eliminated) by another appeal to its explicit definition. Thus, in order to show that the explicit definition that extends $\vdash_1$ to $\vdash_2$ makes for a conservative extension we must show that

> if we introduce the definiendum, only to eliminate it later in a deduction, then we never needed to introduce it in the first place.

but that should be a familiar phrase, it is similar to the definition of harmony (in its second sense) above. This is the sense I shall work with.

> A connective *Con* is legitimately defined when it is defined so that any introduction of it, and subsequent elimination (using the definition of *Con*) can be removed.

For example, suppose that, during a deduction, we reason that

$$\begin{array}{c} \cancel{A} \\ \vdots \\ B \end{array}$$

and then, appealing to the introduction rule for 'if. . . then. . . ' (see 6.1), we introduce $\ulcorner \text{if } A \text{ then } B \urcorner$. Suppose that later in the deduction $\ulcorner \text{if } A \text{ then } B \urcorner$ is used, by appeal to its definition, to legitimate a deduction of B from A. In such a case we did not need to introduce $\ulcorner \text{if } A \text{ then } B \urcorner$ in order to legitimate a deduction of B from A because we already had a deduction of B from A (the deduction we used to introduce $\ulcorner \text{if } A \text{ then } B \urcorner$). More formally

$$\dfrac{\dfrac{\begin{array}{c}\cancel{A}\\ \vdots\\ B\end{array}}{\text{if } A \text{ then } B} \quad \begin{array}{c}\vdots\\ A\end{array}}{B}$$

may be replaced simply by this:[8]

$$\begin{array}{c}\vdots\\ A\\ \vdots\\ B\end{array}$$

and the introduction and subsequent elimination of $\ulcorner \text{if } A \text{ then } B \urcorner$ has been removed.

I conclude from this that harmony in this second sense entails conservativeness and serves as a good condition on the legitimacy of implicit definition. The way that harmony (in the second sense) entails conservativeness is similar to the way (i-iii) entail conservativeness for explicit definitions.[9]

Henceforth when I write of a connective being in harmony I use the term in its second sense.

[8] By tacking the inference of B from $\cancel{A}$ onto the inference of A.

[9] This in itself is of little significance except that if the proofs of conservativeness were very different then we would have a reason to worry that (i-iii) to explicit definitions is not what harmony is to implicit definitions.

Chapter 4

Analyticity and its critics

In this chapter I shall discuss in more detail the objections to analyticity I have already described. I shall discuss and reject objections of Boghossian that there is a fundamental incoherence with the truth-by-definition approach. I shall discuss Quine's objection that there is no fact of the matter about what definitions have been made and reject an attempt at a solution to this objection in terms of inferential roles and practices (I shall not present my own solution to this objection in this chapter, I postulate one in chapter 14. Finally I shall discuss in more detail the objection that there is no coherent theory of implicit definition that does not allow spurious definitions. My response to this final objection is that all the examples used are not in harmony and that the harmony requirement filters out the known problem cases (and that, since harmony entails conservativeness, we have reason to believe there are no problem cases that satisfy the harmony requirement).

4.1 What analyticity does

A theory of analyticity is gets us much of the to way answering this question: why, what and how is a priori knowledge?[1] An analytic sentence is true by its very meaning, knowledge of meaning (in such cases) is an a priori affair.[2] and so we can know that analytic sentences are true simply by understanding them.

[1] If there is such a thing as synthetic a priori knowledge then we must have some idea of what analytic knowledge is to make a coherent distinction.

[2] As the sort of experience we wish to rule out when we say 'a priori knowledge is knowledge without experience' is not the experience of possessing a language.

As I stated in section 2.5 my account of the analytic is in terms of definitions. A sentences is analytic if it can be deduced from the definitions of its terms. Questions of truth/validity, knowledge and meaning of analytic sentences/inferences are answered by emphasising different parts the phrase 'because we defined the relevant terms such that they do':[3]

- Why are the sentences/inferences true/valid? Because we *defined* the relevant terms such that they do.
- Why are the sentences/inferences known? Because *we* defined the relevant terms such that they do.
- Why do the relevant terms have the meaning they do? Because we defined *the relevant terms* such that they do.

The simplicity and intuitiveness of this account is its most attractive feature.

Furthermore it provides a nice link between metaphysical and semantic notions (e.g. truth and meaning) on the one hand, and epistemic and psychological notions (knowledge and information content) on the other. The truth of an analytic sentence is determined (and explained) by its meaning, (the relevant aspect of) its meaning is determined by the definitions of (some of) its terms, the (lack of) information content is explained by the definitions of its terms, and our knowledge of analytic truths is explained by their easily acquired information content.

4.1.1 The lines of criticism

The sort of account of analyticity I have presented has high aims, it is the same sort of account as the original theories of Kant and Leibniz. My account attempts to achieve these by the familiarly contested means of Leibniz, Kant and Carnap (in terms of definitions).

There are two main lines of criticism that have been raised against such accounts:

1. It is often argued that the aims are too high, in particular it is argued that analyticity can explain only the a prioricity of some truths and not why they are true. In other words, it is suggested that it is incoherent that truth can be determined or explained by meaning.

[3] The relevant terms are the terms used in the deduction of the sentence or inference rule in question, so the relevant term in an instance of modus ponens (when worrying about its validity) is 'if... then... '.

2. Quine famously argued against the means by which the aims are to be achieved. Quine attacked the idea that meaning could be determined or explained by definitions.

4.1.2 Some preliminary remarks on definitions

When I come to discuss the epistemology of definitions I shall go into more detail about what a definition is. I take the making of a definition to be an event or collection of events whereby a new term is added to a language. The new term may be

- explicitly defined by some pre-existing construction. For example, we can already express 'unmarried man' and, at some point, there is an *event* where we add a new term 'bachelor' and stipulate that it is to be logically equivalent to 'unmarried man'.

  As Quine suggested, it is rare that a word of English be defined explicitly in this way. When, Quine might ask, was the event where 'bachelor' was defined by pre-existing words 'unmarried' and 'man'? The answer is plausibly never. Indeed, in order for a majority of our words to have been acquired in this way we must have had a comparable expressive power before having obtained a language. That so much innate verbal power should exist is doubtful.

- implicitly defined by accepting some uses of it. The classic example is conjunction. A speaker might not have a word corresponding to conjunction, but he can obtain one. It is obtained by the speaker stipulating that the introduction and elimination rules for conjunction are valid. Or, as suggested by an inferential role theorist, by the speaker using the term in a certain way over a period of time.

An explicit definition can be made only if there already is a term (compound or simple) that expresses what the defined term is to express. However an implicit definition allows us to increase our expressive power, and there are many examples of how this can lead to disaster.[4] It seems that an implicit definition is legitimate (or is genuinely a definition) provided it does

[4] Prior's connective tonk is famous example of this. Here is a simpler example. I might implicitly define the term 'splurg' by accepting that

> people are splurg and only cats are splurg

it follows from this that all people are cats, an empirical and a posteriori sentence if ever there was one.

not entail anything false. If this is true then the project of using definitions as the basis of analyticity collapses.

4.2 Boghossian's criticism

4.2.1 Metaphysical or Epistemic? Metaphysical!

Boghossian [**?**] distinguishes between *metaphysical* and *epistemic* analyticity.

- A sentence is metaphysically analytic when it is true merely in virtue of the meanings of (some of) its terms.
- A sentence is epistemically analytic when a mere grasp of its meaning is sufficient justification for believing it (is true).[5]

Despite Boghossian's objections to the coherence of a dogma of metaphysical analyticity, I claim that the logical truths are analytic in its metaphysical sense.

Boghossian's objection to the metaphysical theory of analyticity is this:

> How could the *mere* fact that S means that p make it the case that S is true? Doesn't it also have to be the case that p?

and

> what is to prevent us from saying that the truth expressed by "Copper is copper" depends in part on a general feature of the way the world is, name that everything is self identical.[6]

Boghossian then discusses and rejects, correctly, a response to this objection on the lines that the meaning of S *makes it the case that p*. Unsurprisingly there are strong objections to this response, for it encapsulates all that is worst about antirealism.

Looking at Boghossian's objection, and trying to make sense of the question 'doesn't it also have to be the case that p', the variable p, presumably, ranges over facts or propositions.[7] The objection appears to me to presume the following picture:

[5] Note this immediately suffers from problems when we ask how it applies to very complicated sentences the meaning of which is beyond our grasp.

[6] This is a quotation from Harman and is extracted from Boghossian's paper.

[7] If p is a sentence variable like 'S' then I cannot understand the quotation as an objection.

> The meaning of a sentence determines what proposition/fact it expresses. A sentence is true when its expressed proposition/fact is true/existent. The truth of a proposition or existence of a fact is mind and meaning independent.

since the truth of a sentence depends on something that is meaning independent, the metaphysical approach to analyticity cannot succeed. This picture may be rejected, in particular by someone who rejects the coherence of facts or propositions.

Let us assume the picture of truth in terms of propositions or facts. Now suppose that the terms of sentence S are such that whatever proposition S expresses, it express a true one. That is, solely in virtue of the meanings of some terms of S, S cannot express a false proposition. Such sentences are analytic. Moreover, they are metaphysically analytic. Metaphysical analyticity is consistent with the view that a sentence is true only because the proposition it expresses is true (or because it expresses some fact). A sentence is true merely because of its meaning when it cannot fail to express a true proposition only because of the meanings of its terms.

For example

> (†) If rice is white then rice is white

is true, it cannot express a false proposition. Which true proposition it expresses exactly depends on what 'rice is white' expresses. However does not (and cannot) express a false proposition because of the meaning of 'if...then...' (and the structure of the sentence). With analyticity seen this way (as it was always intended to be seen) a response to the quotation from Harmon is 'nothing (in the theory of analyticity) is to prevent you from saying it, and to say so is consistent with the metaphysical theory of analyticity'.

I think Boghossian and Harmon's mistake is to read this formulation of the metaphysical thesis literally

> an analytic sentence is one where the meanings of the terms *makes it true.*

As far as I can tell the use of the phrase 'makes it true' in describing the metaphysical thesis of analyticity of Boghossian's origin. No surprise then that he has little trouble refuting it. The correct thesis about metaphysical analyticity is that, a sentence has some terms with certain meanings, and for some sentences the meanings of some of the terms are such that any sentence of the same form is true.[8]

Suppose a new term 'splurg' is defined. We are told that at least all men are splurge, by definition. We know then, that I am splurg (knowing that I am male). We might ask why 'Gabbay is splurg' is true. The answer might be that 'Gabbay is splurg' expresses a proposition/fact that is true/exists. I find this answer unsatisfactory for it neglects to show how 'Gabbay is splurg' came to express *that* proposition/fact. Since the extension of 'splurg' has not been defined it is not clear how it could come to express any particular proposition/fact. Notice that the indeterminacy does not just infect the extension of 'splurg'. It may also be indeterminate whether 'all men are splurg' is intended to be defined for this particular time or over all times and possible situations. That is, if a new male is born, is he also splurg?[9] None of these questions need to be resolved in order to know that 'Gabbay is splurg' is true, but (it seems to me) they must be resolved in order to know what proposition 'Gabbay is splurg' expresses.[10]

Since 'Gabbay is splurg' follows from the information that Gabbay is male, and that 'all males are splurg' is a (metaphysically) analytic truth in the present context,[11] it is true. This simple explanation is not open to one who denies metaphysical analyticity.

Furthermore I find epistemic analyticity unsatisfying, being justified in believing something does not entail its truth. For example:

(†) that man is a man

anyone understanding the meanings of the terms (and knowing that 'that' has a reference) is justified in believing it. However, at least according the theory of indexicals found in [**?**] † might be false. Anyone knowing the

[8]This presumes a notion of form and typing of sentences, I take it that such a presumption is unproblematic.

[9]If there is an answer to this and it and it is 'no' then, at the time of the definition 'all men are splurg' is analytic, a priori, but not necessary.

[10]It is coherent to argue that 'Gabbay is splurg' expresses no proposition, or expresses an indeterminate proposition, or indeterminately expresses a proposition, or expresses a very general proposition, or is supervaluated over a certain extension and anti-extension of propositions, or whatever, provided that 'Gabbay is splurg' comes out as true. In all of these cases some truth conditions of 'Gabbay is splurg' are obtained prior to it expressing any *particular* proposition. This, I think, is enough to make my point. My point is that some sentences can be true only because *whatever proposition they express*, if they come to express one at all, they express a true one, and that this can happen by a stipulation.

[11]For example if the definition of splurg was not made too long ago, for otherwise we might not be sure which men it was being defined over (all men ever, or just the men at the time of definition).

meaning of numbers on a digital clock display is justified in a belief about the time (to a certain degree of accuracy), but that belief could be false.[12] Perhaps justification is meant in its stronger sense, so that

> S is analytic when knowledge K of the meanings of some terms in S is sufficient to provide a justification of the truth of S, where fallibility in this justification derives only from fallibility in knowledge K.[13]

this rules out the examples above, 'that man is a man' and a statement about the time is not epistemically analytic because the justification provided is not infallible enough. I think this will not do, it is too strict. Often justification of a complex proposition derives not simply from the knowledge of the meanings of its terms but also from a deduction of it from sentences which we can already see to be true. A complex proposition (e.g. that there is no greatest prime number) is arguably analytic, but it is not justified merely by knowledge of the meanings of its terms, it is justified also by a deduction. But then, according the epistemic definition, these complex propositions are not analytic. A response to this from a supporter of the epistemic view might be to modify the definition. For generality I assume now that 'analytic' applies to inferences as well as sentences.

1. S is atomically analytic when knowledge K of the meanings of some terms in S is sufficient to provide a justification of a belief that S is true (or that S is valid, if S is an inference), where fallibility in this justification derives only from fallibility in knowledge K.

2. S (a sentence or an inference) is analytic when it is derivable from (or reducible to, in the case of an inference) only atomically analytic sentences by means of atomically analytic inferences.

and then complex propositions and inferences may be epistemically analytic. Notice that the second clause is not epistemic, it requires the possibility of a derivation. I do not doubt this provides a correct account of analyticity, however it assumes a notion of justification which I would like to explain away. If we have a metaphysical approach then the justification for believing an analytic sentence falls out of knowledge of meaning of some terms in S, since the meaning of S determines that S is true.

[12]Perhaps there is a distinction to be made here between meaning and significance, I shall not discuss it any further here.

[13]In other words, the justification is fallible to only to the extent that the knowledge of the meaning S is fallible.

4.2.2 Conclusion

All in all the metaphysical approach, it seems to me, can do everything the epistemic approach can do, all without assuming a notion of justification. Furthermore everything is so much simpler, the corresponding metaphysical definition of analyticity

1. S is atomically analytic when it is true/valid virtue in of the meanings of some of its terms.

2. S is analytic when it is derivable from only atomically analytic sentences by means of atomically analytic inferences.

simplifies greatly. This is because '. . . is true solely in virtue of meanings of some terms of . . . ' is transitive, 'atomically' and the entire second clause is redundant:

> S is analytic when it is true/valid virtue of the meanings of some of its terms.

I conclude that Boghossian's objections to the metaphysical implications of a theory of analyticity are successful only if the theory treats analyticity as a property of propositions. Boghossian's objections have no force if, as I have done, we treat truth as a property of propositions and analyticity as a property of sentences.

4.3 Inferential roles and the criticisms of Quine and Horwich

4.3.1 Two dogmas

I now consider a famous rejection of the analytic-synthetic distinction from Quine.

Boghossian notes well of Quine's famous paper [**?**] that it is not clear whether Quine is arguing that

(NF) No coherent, determinate property is expressed by the predicate 'is analytic'. . . consequently, no coherent expression is expressed by sentences of the form 'S is analytic' and 'S is synthetic'.

(ET) There is a coherent, determinate property expressed by 'is analytic', but, with the exception of those instances that are generated by stipulational mechanisms, it is necessarily uninstantiated.[14]

[14] See [**?**, p340-2].

Either of these theses is damaging to my proposal. Since there is question as to which of these is Quine's is conclusion I shall assume (for the purposes of rejecting the argument) that the conclusion of Quine's argument is

> *If* there is a coherent, determinate property expressed by 'is analytic', *then*, with the exception of those instances that are generated by stipulational mechanisms, it is necessarily uninstantiated.

and argue that show how Quine does not establish this.[15]

Initially Quine writes as if his argumentation does not apply to my project, my project concerns the logical truths, whereas Quine writes that

> the major difficultly lies not in the first class of analytic statements, the logical truths, but rather in the second class which depends on the notion of synonymy.[16]

The first class is my primary concern, the second class of sentences contains analytic truths such as

> All bachelors are unmarried
> A vixen is a female fox.

Quine's arguments are against a claim that these sentences (of the second class) are analytic, not a general claim that there is no notion of analyticity altogether. For example, if an explicit definition has been made in the past then Quine has no problem, for

> ...the definiendum becomes synonymous with the definiens simply because it has been created expressly for the purpose of being synonymous with the definiens. Here we have a really transparent case of synonymy created by definition; would that all species of synonymy were as intelligible.

Perhaps what has confused so many in the past is that Quine suggests later in the paper, after his main arguments, that in the web of belief even the logical truths are subject to empirical revision, no statement is immune to revision, Quine notes that the revision of the law of excluded middle has

[15]Boghossian himself argues that (NF) and (ET) entail a global meaning scepticism akin to that of *Word and object*. He then takes the responses to Quines indeterminacy thesis to apply to the argument of *Two dogmas*.

[16]This is the last sentence of section I in [**?**].

been *proposed* for Quantum mechanics. Whatever is to be made of this comment it is not argued for by the preceding arguments.

The master argument of the paper is that - aside from explicitly defined synonymy and the logical truths – any account of synonymy presumes an account of analyticity and vice versa. For example, dictionary definitions of synonymy are discovered empirically and are subject to revision, furthermore differing interpretations of the linguistic data are available (there is no true dictionary). Quine then notes that Frege's account of analyticity presumes synonymy (or at least, he notes it of a theory he calls 'Frege-analyticity').[17]

The argument against Carnap's account of synonymy is a little more complicated. Carnap divides a language into a basic logic and a set of meaning postulates, if L_n is a particular language then Carnap can definition being 'L_n-analytic' as following from the logic and meaning postulates of L_n. Quine's objection to this is that our language is not an L_n language, to presume such would be to presume the existence of a set of meaning postulates. This presumes a fixed account of synonymy (on which the account of analyticity depends). Effectively Quine is objecting that there is no fact of the matter about which truths we try our hardest not to revise are meaning postulates and which are simply empirical truths that we happen to find impossible to refute.

I think the argument of *Two dogmas of empiricism* is best understood as a challenge to find the meaning facts.

I shall respond to this by finding the meaning facts, I look for them in the realm of definitions. Quine discusses definition briefly in section II and in more detail in *Truth by convention.* Here is a quotation from *Two dogmas.*

> In formal and informal work alike, thus, we find that definition – except in the extreme case of explicitly conventional introduction of new notations - hinges on prior relations of synonymy. Recognising that the notion of definition does not hold the key to synonymy and analyticity, let us look further into synonymy and say no more of definition.

I take it that this comment, if it is true to the remainder of the text, renders Quine's paper outdated with respect to its discussions of definition. The theory of definitions and other meaning fixing events has advanced significantly to render such a dismissal unwarranted now. Nevertheless, as we shall see, Quine's point is easily extended to the more modern theories of definition.

[17]The account is that analytic statements are derived by substituting synonymous expressions into logical truths.

Modern worries about definitions lie in whether conditions can be given for the more subtle types of definition that Quine does not discuss, implicit definitions, to ensure the definition really defines something meaningful (e.g. Prior's connective 'tonk').

4.3.2 Inferential roles

Quine's arguments, although not originally applied to logic, may be extended. We may ask what the meaning facts are that give the logical constants their unrevisable status. Are they simply so deep in our web of belief that they are hard to refute empirically to the point of appearing paradigms of analyticity? An attempt to resolve this worry, at least for logical truths, is to enrich the account of definitions.

What is apparent from our reasoning is that some terms are used by us such that they fulfil an inferential role. For example, conjunction fulfils a role of allowing each conjunct to be inferred from it. Perhaps, the terms are used *in order to fulfil these roles.*[18] In this case terms may be defined such that they fulfil inferential roles. It is claimed that such definition, provided that there is a role to be discerned, is as unproblematic as the extreme cases of explicit definition that Quine chooses to ignore. Such definitions are commonly called implicit definitions.

For example the word 'and' may be implicitly defined by using it to fulfil the following inferential role:

From the premises A, B infer that $C(A, B)$

From the premise $C(A, B)$ infer the conclusions A, B

where C stands in for a concept or a word or whatever an inferential role is a role of.

A sentence is analytic, on an inferential role theory, when it is derivable from the inference patterns that are its inferential roles. For example, the inferential role of 'if...then...' includes the inferences modus ponens and conditional proof and using them we may infer $\ulcorner$if A then $B\urcorner$. Consequently $\ulcorner$if A then $A\urcorner$ is analytic and a priori (since knowledge of the inferential roles is a priori, after all, they are our implicit definitions), and also necessary (as to use the terms otherwise, so that $\ulcorner$if A then $A\urcorner$ is not derivable, is to use them not in those inferential roles and thus so they have different meanings).

The two worries with inferential role theories can be summed up by these two questions:

[18] Or to label concepts that fulfil these roles.

1. Which rules are being followed anyway? This is the Kripke-Quine worry about indeterminacy.

2. Which rule following is meaning constitutive? There are some rules that appear to have no meaning (e.g. Prior's rules for tonk). Perhaps an answer to this cannot be given without presupposing an account of analyticity.

4.3.3 The Quinean worry of determinacy

A serious problem for an inferential role theorist is to give an account of what fact of the matter there is that determines precisely what role we are fulfilling with our inferences. For example we may see a speaker reasoning, apparently in accordance with these rules, for 'and'

> From the premises A, B infer the conclusion $\ulcorner A$ and $B \urcorner$
> From the premise $\ulcorner A$ and $B \urcorner$ infer the conclusions A, B

it seems that 'and' is used to fulfil an inferential role and is thereby defined to have a certain meaning. The same speaker also reasons as follows

> From the premise that A infer 'God knows that A'

> From the premise 'God knows that A' infer that A

the speaker may make these inferences on the basis of a religious experience or through religious education or simply because he finds a belief in a supreme omniscient being compelling. Intuitively 'God knows that...' is not fulfilling any inferential role. Certainly these inferences do not fix the *meaning* of the expression 'God knows that' such that the inferences above are valid.[19] One might argue that the difference (between 'God knows that...' and '...and...') lies in other beliefs the speaker might have about the nature God. 'God' comes as part of a large religious doctrine, whereas 'and' is a single logical connective which is thought of only in terms of the inference patters above. However a speaker may have many beliefs about the nature (e.g. the logical properties) about conjunction which are part of large philosophical doctrines (e.g. the speaker may believe, among other things, that 'and' is a paradigm of an implicitly defined connective).

Any data on what rules a speaker is following with regard to a certain term underdetermines the exact nature of the rule (as the data will not cover

[19] A proof of divine existence should not come so easily.

every case). If truth conditions and meaning depend on the rule then the data underdetermines the truth conditions and hence the meaning of the term. The term could be applied to an infinity of cases whereas the speaker will have used it in only a finite number of cases. Looking at the data on the speakers' linguistic practices there is not enough to determine what rule exactly the speaker is following. For example every speaker (of a certain class of languages) living at around now uses connectives, denote them by $\oplus$ and $\otimes$, that follow these rules

$$\left.\begin{array}{lll} \dfrac{A \quad B}{A \oplus B}\,\oplus I & \dfrac{A \oplus B}{A}\,\oplus E & \dfrac{A \oplus A}{B}\,\oplus E \\[3ex] \dfrac{\begin{array}{c}A\\ \vdots\\ B\end{array}}{A \otimes B}\,\otimes I & \dfrac{A \quad A \otimes B}{A}\,\otimes E & \end{array}\right\} \text{before 1000 A.D.}$$

$$\left.\begin{array}{lll} \dfrac{A \quad B}{A \otimes B}\,\otimes I & \dfrac{A \otimes B}{A}\,\otimes E & \dfrac{A \otimes A}{B}\,\otimes E \\[3ex] \dfrac{\begin{array}{c}A\\ \vdots\\ B\end{array}}{A \oplus B}\,\oplus I & \dfrac{A \quad A \oplus B}{A}\,\oplus E & \end{array}\right\} \text{after 1000 A.D.}$$

I take it that nobody has spanned the millennium, so everyone has been following these rules. That is, everyone has been following a rule that is extensionally identical to this rule (though not necessarily extensionally equivalent). This of course does not mean that people have been following those particular rules. However, if all there is to following a rule is to accept some inferences as true (i.e. to act such that certain data holds) then there is no difference between following one rule and following one extensionally identical rule. Put in a different way, the inferential roles of 'and' and 'if. . . then. . . ' are no different than of $\otimes$ and $\oplus$.

Compare with Quine's argument for the indeterminacy of reference. Quine takes meaning to be no more than the providing of a translation manual, by a scientist of the topic. Since no scientist can ever collect enough data to determine his theory, translation hence meaning is underdetermined. Since there is no more to meaning (for Quine) than providing a scientific theory of it, the underdeterminacy becomes indeterminacy. Similarly for Kripke and his connective Quus. The assumption is that the meaning of a term like + consists only in using it in a certain way, following a rule. But the actual use always underdetermines the extension, since actual use is all there is to *following* a rule, the underdeterminacy becomes indeterminacy.

A theory of meaning or truth conditions that entails some indeterminacy is not necessarily false. There is reason to suppose that there is some indeterminacy in meaning (e.g. in vague cases). However the extent of the indeterminacy outlined above is I believe severe. In general, any theory that requires meaning or truth conditions to be determined by behaviour suffers from a similar indeterminacy.

Consider the following cases. It seems that an inferential role theorist, having only behaviour to observe, cannot discern errors.

- A speaker may use a term with one meaning but following the wrong rule. For example a speaker may mean 'and' but eliminate it as if it were 'or'.[20]
- A speaker may use a term incorrectly with a certain meaning. For example, committing an arithmetical error.
- A speaker may use a term with a certain meaning without knowing any rule to apply to it. For example when learning new terms.

The third perhaps can be explained in terms of intentions, a term can have meaning (or truth conditions or reference) if we hear it from someone else and use it with the intention of it meaning the same. However the first two are not so easily accounted for.

It has yet to be explained what fact there is that distinguishes a deviation from a certain rule from the following of a different rule. If inferential role is determined by actual inferences and inferential role is all there is to the truth conditions (or meanings) of the logical connectives then there is no fact that distinguishes an error from a different inferential role.[21]

No doubt a principle of charity and radical interpretation will, to some extent, reduce the indeterminacy in what rules satisfy the data obtained to date. However this will be of little help with projections into the future, many rules which differ greatly from each other meet the criterion of the

[20] In this case the elimination rule for disjunction is valid for conjunction, the speaker will not be acting inconsistently.

[21] Not all errors cause inconsistencies, appeal to consistency is not sufficient here. Appeals to consistency (e.g. to rule out tonk) seem somewhat circular to me in the case of logical connectives as the system we get from defining the connectives is what we use to determine consistency (unless we allow for the power of logic to be present already in some sense, which I argue it is above, in which case we have something better than behaviour to determine the truth conditions of our connectives).

radical interpretation. We could supply a theory of projectibility of rules (like providing a theory of what predicate are projectible in inductions) but, as the Quine of *Two Dogmas* would be quick to retort, this is tantamount to assuming analyticity beforehand. On what basis are these decisions made?[22]

It is too much implicitness to have the implicit definitions made implicitly by a speaker's actual inferences. This puts too much of a priority on a speaker's behaviour.[23]

A possible response

The problem of indeterminacy need not be framed against a functional role theorist. Even if the implicit definition is made explicitly there is argument to be made that it does not define a term with unique meaning or truth conditions. There may be many candidates that satisfy the inference rules. Here is another example: although conjunction (...and...) satisfies the usual introduction and elimination rules, so does something more complicated like '...at the same time as...', or even '...and...while 1+1=2'.

A natural answer is to take the 'simplest' candidate, or class of candidates (treating members of such a class as synonymous). We can give a more rigorous account of simplicity here, one meaning a is simpler than another meaning b when, if someone has the capacity to obtain b then he has the capacity to obtain a.[24] In appendix 14.3 I suggest conditions that are the minimum required for a agent to be able to reason (these conditions allow the implicit definition of the logical connectives), being able to handle arithmetic, identity and time is fundamental to thought, but not the minimum. Thus '...and...' is simpler than '...at the same time as...'. The usual meanings are clearly candidates for the logical connectives (implicitly defined by our introduction and elimination rules). This handles the uniqueness problem and the indeterminacy.

This response is not open to a inferential role theorist, what functional role a connective is filling depends on how the speaker is using it. The sense of simplicity for inferential roles would be that a is simpler than b if whatever capacity a speaker requires to have a connective filling role b also fills role

[22] I think an antirealist can provide an answer to this: the conservative rules are projectible and they are the ones that yield analyticity, regardless of what the speaker thinks he is doing. I have no objection to this other than that this makes analyticity a mere epiphenomenon of the way a speaker happens to be using his language. Perhaps such a conclusion is part of the point of antirealism, so much the worse for an antirealist I think.

[23] Either in his public or mental life, not all errors or sentences are spoken to others (or spoken at all).

[24] 'Obtain' here means 'express terms with the meaning of'.

a. But given the date, the usual 'and' and $\otimes$ have the same inferential role, neither is simpler than the other. Moreover, since the term has only been used in a finite number of cases, at any time there is an infinity of different rules all fulfilling the same role just for those cases, no one of them can be said to be simpler that any of the others.

A response to the indeterminacy in terms of simplicity is applicable if the implicit definition is made explicitly, for then we can ask what goes on in the speakers mind for him to make such a definition, we can ask what his intentions were. The answer 'the simplest thing' becomes more reasonable. I see more future therefore in ground analyticity not in implicit implicit definitions (is the inferential role theory does) but in explicit implicit definitions (i.e. theorise that there was a time and a place where the inference rules were used, explicitly, as deifnitions). I postulate an account of how this might work in chapter 14.

4.3.4 The worry of reliability

The second question (of section 4.3.2) is asked bearing in mind that not just any inferential roles yield a meaning, some roles are inconsistent (e.g. tonk). It is natrual to say that such roles do not make for legitimate implicit definitions, but this requires a condition on legitimacy that does not assume an account of analyticity.

The worry behind the question comes from the fact that an implicit definition defines meaning or at least the truth conditions of a term by stipulating sentences containing that term as true, or certain inferences involving that term as valid. Much of the problems with implicit definitions centre around this stipulation, e.g. it is not such an easy thing to stipulate that something is *true*, especially if it already has some degree of substance.

A natural scientist of the late 17th-century may have introduced the term 'phlogiston' by implicit definition

> Phlogiston is released when combustion occurs.

this implicit definition requires a great deal of other beliefs about the nature of combustion to understand it (as that historical theory). The meaning of 'phlogiston' is then dependent on what we already take to be true. With a modern theory of combustion behind us we cannot understand 'phlogiston', by the implicit definition above, to mean as the 17th-century chemist did. As modern chemists we may understand 'phlogiston' by the implicit definition above with the additional clause that the definition is to be taken with certain 17th-century chemistry. Alternatively we may interpret the 17th-century term ‘phlogiston’ in our own theory and refer to it with the same syntax.

Since we now consider phlogiston theory to be false, this kind of example implies the following about implicit definitions.

1. An implicitly defined term might change its meaning as our background beliefs change.

2. We may later discover that a sentence with the same truth conditions to be false (e.g. Lavoisier observed ‘phlogiston exists’ is false).

3. If the background beliefs are false then there may be no, or insubstantial, meaning to the implicitly defined term.

An example of the third is the following definition of the term ‘peaceday’

> Peaceday was when I stopped beating my wife.

The famous, false (in my case) presupposition stops this from being a successful implicit definition, for the sentence is not true and stipulating otherwise will not change this. We may avoid this by applying the Ramsey technique and having the following implicit definition:

> If there was a day I stopped beating my wife then Peaceday was when I stopped beating my wife.

which can be taken as true (even if the conditional is not material) without any fear of falsity.

This is adequate for cases where a defining clause is expressed already in the implicit definition (here ‘whatever day I stopped beating my wife’) for then we may place it in a conditional with an appropriate antecedent (e.g. ‘if there is a day I stopped beating my wife. . . ’). However, it is not clear how to do this for logical connectives, where we cannot assume that a defining clause is expressible to be put in the antecedent. Further, it would not do, in giving an account of how we come to define ‘if. . . then. . . ’, to require that we must already have a conditional to make the implicit definition.

Note that the implicit definitions above involve the implicit definition of a term through the use of other highly substantive terms. For example ‘peaceday’ and ‘phlogiston’ were defined using the terms ‘beating’, ‘wife’ and ‘combustion’. These terms bore substantive background beliefs, e.g. beliefs about the basic composition of substances in the case of phlogiston.

In the case of peaceday these background beliefs may be required even to understand the definition.

Thus we cannot be sure that an implicitly defined term has the meaning we wished, is true, or even means anything at all. This is an unpleasant setback to the program of defining and justifying logical connectives and inferences by means of implicit definitions (as is shown by Prior's connective tonk).

We encounter similar problems if we expect to obtain any significant new knowledge from implicit definitions. An implicitly defined term, ultimately, relies on a stipulation (the definition is the stipulation). As we have seen above not just any stipulation is legitimate.[25] But, as the examples above suggest, it requires a significant amount of knowledge to know that an implicit definition is legitimate. Indeed, it seems to require as much knowledge as we 'gain' from the definition. Further, in many cases, the knowledge of legitimacy is an a posteriori affair; for example we later learned that the background beliefs behind phlogiston theory were not true.

The charge is that implicit definition is not a reliable mechanism for defining terms. We cannot guarantee that the sentence we take to be true, or inference we take to be valid, actually is true or valid. The implicit definition is subject to revision and change as we learn more and make new discoveries. If this is the case, then we should not expect anything derived from implicit definitions to be a priori or even true.

My suggested response

The best response, I think, is to show how the harmony (normalisation) requirement filters out all the bad cases in a non ad hoc way.

It should be clear that implicit definitions made within a revisable framework are themselves subject to revision and are no more a priori than the framework itself; this is true of any term or sentence that requires the framework to be understood. Further, an implicit definition that itself adds to the framework is no more than an unjustified stipulation and will not add to our knowledge even if it happens to be true.

For an implicit definition to have the a priori nature we require we must ensure that

[25] Horwich has another good example, the implicit definition of f such that

snow is green and moon is f

is true, is not a legitimate definition.

1. it is made from within a true (if it contains any assumptions) and unrevisable framework

2. the implicit definition adds no new information that does not involve the defined term.

But how is it best to understand the phrase 'adds nothing new'? An implicit definition adds nothing new if we can always eliminate it without affecting our reasoning or knowledge. That is, an implicit definition of t is eliminable when it is not necessary for achieving knowledge of any proposition that does not involve t.[26] But this is exactly the requirement of harmony (see 3.3.1), if our deductions normalise (the connective is in harmony) then the new connective is eliminable. Also note the identity between these ideas and much of the discussion of conservativeness in 3.3.1.

[26] The definition in Horwich's counterexample is not eliminable, for we may use it do deduce something new: snow is green. Further the implicit definition is necessary for the deduction of this.

Chapter 5

Other responses to worries about reliability

The theories discussed here are not primarily responses to the Quinean sceptical worry outlined in section 4.3.3, although finding what response they would give is not difficult. I discuss two attepts at responding to the worry of section 4.3.4.

One form of response, by Peacocke, is to use a semantics to discern a genuine inferential role from a non genuine one. For example a semantics does not necessarily contain an element 'God', but it does contain truth value assignments. Thus conjunction obtains meaning by fulfilling a role underpinned by truth value assignments whereas 'God knows that...' has no primitive semantic element to give it its meaning (so as to refer to a supreme being), so it is not analytic.

Another response, by Dummett, is to give syntactic considerations on what rules are legitimate (being an antirealist, semantic considerations do not come so easily to Dummett).

I shall reject Peacocke's account and not object greatly to Dummett's account. I discuss the two accounts here for two reasons. Firstly it is interesting to see how both accounts, coming from different sides of the realism vs antirealism debate both rely on something like the harmony requirement to work. Secondly I think a rejection of Peacocke's strategy helps motivate (by a the-alternatives-fail argument) the strategy I employ in chapter 14.

5.1 Peacocke's realism

Peacocke's realist approach may be found in [**?**]. Peacocke argues that

1. The famous rules for the logical connectives are primitively obvious
2. Finding such inference rules primitively obvious is partially constitutive of understanding it [the logical connective].
3. The semantic value of a logical constant is the function on semantic values on which the constant operates which ensures that
   (a) the principles the speaker finds primitively obvious for it are truth preserving
   (b) any maximal set of limiting principles for the constant are truth preserving.

The limiting principles clause is to allow Peacocke to justify the existential elimination rule which is certainly not primitively obvious. Peacocke uses the notion of limiting principles to validate elimination rules for logical constants for which only the introduction rules are primitively obvious, in particular the existential quantifier.[1] The existential quantifier may be introduced like this:

$$\frac{Fa}{\exists x Fx}$$

and a speaker may find only rules of this kind primitively obvious. Basically, the limiting principle for $\exists$ is an elimination rule that is validated by the strongest semantic assignment to $\exists$ that validates the introduction rules. Peacocke claims that the classical interpretation of the existential quantifier is the only such principle and so a maximal set of limiting principles contains only it.

I shall discuss now the idea behind the limiting principle clause. The idea is this, take the strongest semantic assignment for the introduction rules and then find an elimination rule for which it is also the strongest assignment. Any speaker is then justified in using this elimination rule. Why the strongest semantic assignment? Let the *canonical grounds* (Peacocke's terminology) be the premises on which the connective is introduced, a speaker may then realise that his introduction rules exhaust the *range of canonical grounds* for the connective and that there are no others. If the introduction rules do exhaust the canonical grounds for the connective then it should be assigned the strongest semantics. Why? Peacocke does not say,

[1] In my discussion of first order logic I shall use an elimination for the existential quantifier which far simpler than Peackocke's and is, I think, obvious.

neither does he say where the 'range' of canonical grounds comes from aside from the introduction rules supplied for the connective.

Perhaps the range of canonical grounds derives from the semantics and that the actual canonical grounds derives from the syntax. That is, the semantics tells us in what cases we may introduce an certain connective, and if these are exactly the cases considered by the (premises of the) introduction rules for the connective then the connective should be assigned the strongest semantic value satisfying the introduction rules and from this we may justify an elimination rule (even if no speaker can work out what that rule is exactly). This works only if the connective in question has already been assigned a semantics, to assume this begs the question entirely for we seek to answer how each connective obtains its semantic value, we cannot simply assume that it already has one.

Perhaps the range of canonical grounds is not obtained from the semantics, then I do not see how it could differ from the premises of the introduction rules, in which case the introduction rules trivially exhaust the range of canonical grounds for a connective. Peacocke's suggestion might be that if certain introduction rules are *all* the introduction rules then the connective should be assigned the strongest semantics so as to ensure that no other potential introduction rules are validated (so that the introduction rules are the only valid rules for introducing the connective). But then, in the case of the existential quantifier, I do not see how this justifies the classical elimination rule. The strongest semantics that validate the existential introduction rules is not the classical use. Consider what I shall call the *independent existential quantifier*, this is where 'something' *refers* to something in the domain rather than quantifying.[2]

A semantics for the independent existential quantifier is easily obtained. Let a model M consist of a domain D and an interpretation I of the predicates and constants,[3] Let v be a valuation that assigns elements of D to variables as with usual semantics and Formulae are be satisfied by pairs consisting of a valuation and a set of variables. Also, as in the usual semantics for classical logic, formulae may be satisfied by a valuation alone (not in

[2]For example, Peter might see Jane in a state of extreme anger and say 'someone is not a happy bunny'. Or perhaps the reading of 'everyone loves someone' where it is the same 'someone' being loved. This reading may not, as usually characterised, be a matter of scope, but in fact be a matter of the word 'someone' having a particular reference. Whether there really is such a use of the term in English is beside the point for a semantics can be defined for it, if English does not possess an independent existential quantifier, so much the worse for Peacocke as his thesis, it seems to me, entails that it (or something even stronger, if there is such a thing) is the correct semantics for the existential quantifier.

[3]I assigns sets of n-tuples of elements of D to n-ary predicates etc.

pair). Let U be a set of variables

- $v(\perp) \neq T$
- $\langle U, v\rangle(P(t_1 \ldots t_n)) = T$ iff every valuation v' that agrees with v on everything except perhaps on what is assigned to the variables of U is such that $v'(t_1) \ldots v'(t_n)$ belongs to P^I.[4].
- $\langle U, v\rangle(B \wedge C) = T$ iff $\langle U, v\rangle(B)$ and $\langle U, v\rangle(C)$ are true
- $\langle U, v\rangle(B \vee C) = T$ iff $v'(B)$ or $v'(C)$ is true for every valuation v' that agrees with v on everything accept perhaps the members of U.
- $\langle U, v\rangle(\sim B) = T$ iff $v'(B)$ is not true for every valuation v' that agrees with v on everything accept perhaps the members of U.
- $\langle U, v\rangle(\forall x B) = T$ iff $\langle U \cup \{x\}, v\rangle(B)$ is true.
- $\langle U, v\rangle(\exists x B) = T$ iff $\langle U, v'\rangle(B)$ is true for some valuation v' that agrees with v on everything except perhaps on x.

$\Gamma \vDash A$ when for any model, $v(A) = T$ whenever $v(\Gamma) = T$.[5] Also we may show, by induction on the degree of A, that $\langle U, v\rangle(A) = T$ iff $v'(A) = T$ for every valuation v' that agrees with v on everything except perhaps the variables in U.

In this semantics the 'value' of the existential quantifier is an element of the domain that does not depend on any universal quantifiers that contain it (the existential quantifier) in their scope. The usual inference rules are all validated by this semantics.[6] In this semantics, $\forall x \exists y A$ is interchangeable (in any formula) with $\exists y \forall x A$,[7] but also $\exists x A$ is interchangeable with $\sim\forall x \sim A$. So the semantic assignment above is clearly stronger than the classical semantic

[4]Note that $v(B)$ has the same value as $\langle \emptyset, v\rangle(B)$

[5]This semantics does not involve truth value gaps, an induction on the degree of A shows that $v(A) \neq T$ iff $v(\sim A) = T$. Thus if U is empty then the semantic definitions above are equivalent to the usual classical ones.

[6]By induction on the length of a deduction $\Gamma \vdash A$ we must show that if $v(\Gamma) = T$ then $v(A) = T$ is true. For example if the last step in the deduction is $\sim I$ then by induction hypothesis if $v(\Gamma \cup \{B\}) = T$ then $v(\perp) = T$, thus $v(\Gamma \cup \{B\}) \neq T$ $v(\sim B) = T$.

[7]To see this we need only verify that $\langle U, v\rangle \forall x \exists y A = T$ iff $\langle U, v\rangle \exists y \forall x A = T$:

$\langle U, v\rangle(\forall x \exists y A) = T$ iff $\langle U \cup \{x\}, v\rangle(\exists y A) = T$ iff $\langle U \cup \{x\}, v'\rangle(A) = T$ iff $\langle U, v'\rangle(\forall x A) = T$ iff $\langle U, v\rangle(\exists y \forall x A) = T$

assignment of the existential quantifier, and supports this rule in addition to the usual rules:

$$\frac{\forall x \exists y A}{\exists y \forall x A}\ \exists P$$

If talk involving limiting principles is necessary to Peacocke's thesis then he does not obtain the logic he wishes as classical first order logic is not the logic that satisfies his conditions (it is not based on the strongest semantics for which the introduction rules are sound).

Problems with limiting principles aside, I think Peacocke's approach suffers from a problem which arises from his reliance on semantics. I think the question of what a correct semantics is like should be at least in part answered by the question of what rules govern the logical connectives. Since a major part of the semantics is derived from the logic, I claim that it is begging the question to assume a certain semantics when explaining why certain logical inferences are valid.[8]

In Peacocke's discussion of classical negation. The negation of A is, according to Peacocke, the weakest 'condition' (Peacocke's terminology) incompatible with A.[9] This then justifies the famous inference rules for negation. Furthermore Peackocke claims, without argument, that the double negation elimination rule follows from these considerations, i.e. that A follows from not-not-A. It follows from this that A and not-not-A are equivalent. But without the double negation elimination rule not-not-A is weaker than A, so it seems that without the double negation elimination rule negation is weaker. So if not-not-A is the weakest condition incompatible with not-A we should not allow it to be as strong as A if we can avoid it. In other words we should not accept double negation elimination.

The problem inherent in the discussion above is that there is a shift from a semantic sense of weakness to a proof theoretic sense of weakness. Proof theoretically speaking, classical logic is stronger than intuitionistic logic as classical logic has more theorems. However semantically speaking, intuitionistic logic is stronger.[10] Peacocke clearly intends 'weakest' to be

[8]Perhaps this does not apply to Peacocke if he rejects my claim that syntax is prior to semantics. If this is so and such a position is tenable, then Peacocke does not beg the question, although I cannot adopt his theory in view of my belief that syntax comes first.

[9]On p164 Peacocke has to change his definition of a limiting principle to the weakest semantic assignment, why the limiting principle for negation requires the weakest assignment whereas for the existential quantifier it requires the strongest assignment is not explained.

[10]Take any semantics for intuitionistic logic, the conditions by which $\neg A$ is true entail the conditions under which A is *not* true, the 'not' in italics is of a classical meta-language. There are well known problems with trying to characterise this 'not' in the object language

in its semantic sense. But what is the semantics? Not every semantics allows for an easy addition of a classical negation, there are semantics for intuitionistic logic where the weakest condition incompatible with A does not support the classical rules. Peacocke seems to have assumed a truth functional semantics for his entire discussion.

Peacocke responds to a general objection that the weakest condition incompatible with A need not validate the rules of classical negation, he assumes that anyone levelling such an objection is an intuitionist. His response is that an intuitionist means something different by 'incompatible' from a classicist. Apparently, in order to be an intuitionist one must be a verificationist, Peacocke claims that an intuitionist means by incompatibility that 'the supposition that A and B are both verified leads to absurdity'.[11]

Even if Peacocke is correct that an intuitionist must mean something different by such terms (which he is not), I still do not see why negation must be classical, unless Peacocke has assumed that there is a correct semantics and that it is classical.

It becomes more apparent that Peacocke has made such an assumption in his discussion of *tonk*. Let us denote tonk by *tonk* and the rules for tonk allow that $A\, tonk\, B$ follows from A, and B follows from $A\, tonk\, B$. Any logic containing tonk either has no theorems or no non-theorems, the challenge posed by tonk is to explain why it cannot be defined into existence. Peacocke disallows tonk as a genuine logical connective on semantic grounds:

> The semantical objection to *tonk* is that there is no binary function on truth values [of the sentences *tonk* connects] which validates both its introduction and elimination rules. (p167)

as it stands this is fine objection to *tonk* for it shows that no semantics could support it.[12] But this constraint does also rules out connectives like the existential quantifier, the existential quantifier cannot be interpreted as a truth function on the truth values of any of its subformulae.

Perhaps I am reading his semantical objection too literally. Earlier he formulates the semantic constraint whereby a new connective with certain laws is legitimately introduced when '*there is* a semantical value for it which

(the two negations collapse into one classical negation).

[11] This perhaps is a common supposition, although intuitionistic logic is often motivated with an appeal to verification or grounds for assertion, I see no reason that an intuitionist necessarily be a verificationist.

[12] On the condition that the semantics allows all four combinations of truth values for any two atomic formulae. If, in the semantics, no proposition is ever false, then tonk is unproblematic.

makes those laws necessarily truth-preserving'. Here he refers to 'semantic value' which need not be a truth function but an appropriate element of the semantics. Peacocke then claims that this semantic constraint ensures that the addition of new logical constant (connective) is a conservative extension.

- Suppose we have a strict conditional $\rightarrow$ with the introduction an elimination rules of intuitionistic logic, we do not wish $[(A \rightarrow B) \rightarrow A] \rightarrow A$ to be necessarily true as $\rightarrow$ is the strict conditional (the conditional of entailment). We then add the classical conditional $\supset$ with its familiar inference rules (which allow the deduction of $[(A \supset B) \supset A] \supset A$). Famously there is a function on truth values that validates $\supset$. The result of this addition is that the strict conditional $\rightarrow$ collapses into the classical conditional $\supset$: for example, $[(A \rightarrow B) \rightarrow A] \rightarrow A$ becomes deducible. Something has gone wrong, this should not happen for the addition of the classical conditional allows us to prove things that are not true of the strict conditional. But Peacocke's requirement is met for the classical conditional, there is a function on truth values that validates its rules, more so, there is a function on truth values that is completely characterised by the classical conditional. Not only should nothing be wrong but according to Peacocke the addition of $\supset$ should be a conservative addition (p168 for a 'theorem' to this effect).

- Let us define the connective Ex such that

$$\frac{A[x/t]}{ExA}\ EI \qquad \frac{ExA}{A[x/a]}\ EE \quad \text{where } a \text{ is a fresh constant}^{13}$$

  Ex is the independent existential quantifier. Adding Ex to a first order logic results in the existential quantifier collapsing into the independent existential quantifier. Since there are fewer restrictions on Ex and the introduction rules for $\exists x$ and Ex are the same, we can turn instances of $\exists x$ into Ex, profit from the weaker restrictions and then turn Ex back into $\exists x$.[14]

[13] a is fresh when a does not appear in any assumptions on which ExA depends, a does not occur in any premises of the deduction as a whole, and a is introduced by applications of EE only with the premise ExA.

[14] Here is a sample deduction:

$$\cfrac{\cfrac{\cfrac{\cfrac{\cfrac{\forall x\exists yFxy}{\exists yFxy}\ \forall E \qquad \cfrac{\cancel{Fxc}^1}{EyFxy}\ EI}{EyFxy}\ \exists E(1)}{Fxa}\ EE}{\forall xFxa}\ \forall I}{\exists y\forall xFxy}\ EI$$

> Since no collapse occurs in natural language we must show what is wrong with such a simple addition. It is not enough to appeal to the nonexistence of any semantic valuations to rule it out, I give a semantics for Ex above.

These two examples cast doubt on the claim that Peacocke's semantic constraint does ensure a conservative extension, or that it ensures an appropriate logic.[15] Certainly if his constraint is what I think it is, then it does not ensure a conservative extension, neither is it in general applicable (e.g. to the existential quantifier).

Perhaps what has gone wrong is that I have not interpreted Peacocke's semantic constraint properly, perhaps Peacocke's semantic constraint really requires that the semantics should already be able to interpret the additional connective. In this case the addition will be conservative and Peacocke's 'proof' that it ensures a conservative extension really is a proof. However what connectives and rules a semantics is able to interpret depends on the semantics. An intutionistic semantics cannot interpret classical negation (hence the collapse when one is added), and many classical semantics (e.g. the usual truth functional semantics) have no way of interpreting a non-truth functional connective except as a truth function. Peacocke's semantic objection to *tonk* either requires a semantics to be fixed prior to what rules and connective we have (I say this begs the question) or it does not work.

Maybe the semantic constraint could be cashed out this way. If logic L is sound and complete for semantics S and S', then if L' (an extension of L) is also (sound and) complete for S', L' is a conservative extension of L. To see this suppose that A is in the language of L and is provable in L', then it is true in S', but since L is complete for S' it is provable also in L. But considerations of conservative extensions are not enough, by itself tonk is not inconsistent and has a semantics.[16] Negation is not a conservative extension over tonk, but that does not rule negation out as illegitimate. Certainly if negation is necessarily added *before* tonk is added then tonk is illegitimate, but why that ordering is so special is yet to be explained.

The problem is the reliance on semantics, either we must assume a particular semantics and beg the question, or we cannot provide adequate con-

[15] A logic with only an existential quantifier that allows 'someone is a parent of everyone' to be deducible from 'everyone is a child of someone' is not appropriate, the independent existential quantifier allows this and we must show how its addition so as to collapse the existential quantifier is not legitimate.

[16] The models of tonk semantics have no interpretation of $\perp$ and assign T to every atomic formula. Tonk then may be interpreted as the truth function 'and' or 'or'.

straints on which connectives are legitimate.

5.2 Dummett's antirealism

Dummett rejects the idea that the meaning of a term is given in terms of truth conditions. The basic objection is that such a theory requires us to know the meaning of some sentences the truth conditions of which are beyond our ken.

Dummett's account of meaning is given in terms of verification. The meaning of a sentence is given by the means of verifying it. Conjunction obtains its meaning by its verification conditions. The inference rules for conjunction tell us that a verification of $\ulcorner A$ and $B\urcorner$ consists of a verification of A and a verification of B. In Dummett's language, what we know when we know the meaning of conjunction is that exactly the verifications (or warrants) for (believing/asserting etc.) A and B are the verifications (or warrants) for (believing/asserting etc.) $\ulcorner A$ and $B\urcorner$. Also for 'if…then…': any verification (or warrant) for inferring B from A is warrant (or verification) for (believing/asserting etc.) $\ulcorner \text{if}\, A\, \text{then}\, B\urcorner$.

The intuition for the replacement of truth conditions with verifications derives from an analogy with mathematics and proofs. Dummett rejects that any mathematical statement is to be understood in terms of truth conditions (on the grounds that the truth conditions may be too complicated, or inaccessible in the case of undecidable statements, for a speaker to grasp). A mathematical statement is understood in terms of what is involved in proving it. The understanding of the logical constants derives from the constructive semantics for mathematics (on which intuitionistic logic basis itself), where mathematical statements are understood in terms of constructions (proofs) rather than truth. The non-mathematical analogue of a proof for Dummett is verification or perhaps more generally warrant.

A fundamental worry for Dummett is that the account of the meanings of the logical connectives in terms of verification seems circular. Consider the explanation of the conditional, a (verified/warranted) conditional allows us to turn an arbitrary verification/warrant of its premise into a verification/warrant of its conclusion. The seems to entail that

> we must, in some sense, be able to survey or grasp some totality of constructions which will include all possible proofs of a given statement.[17]

[17] [**?**, p390]

The problem arises because the addition of the logical connectives, together with their inference rules, has enhanced what it is to be a proof (verification). If we introduce the conditional such that a verification $\ulcorner \text{if } A \text{ then } B \urcorner$ involves a verification P of the consequent from (an assumption of the antecedent), then, what counts as a verification is expanded so that P may involve conditional reasoning.[18] But this seems circular, for the meaning of the conditional appears to depend on a notion of a verification that itself depends on the introduction of a conditional. What constitutes a verification (a proof) depends on what terms are introduced (what they mean), and (the meanings of) what terms are introduced depends on what constitutes a verification (or proof). Not all circularities are vicious, apparent circularity in many conditions may be removed by reformulating the conditions in terms of a recursive condition. Dummett eliminates the circularity using standard methodology: to structure sentences into a hierarchy of increasing complexity and define a verification (a proof) recursively on the levels. For example

- there are atomic verifications
- a verification of $\ulcorner A$ and $B \urcorner$ or $\ulcorner \text{if } A \text{ then } B \urcorner$ etc. depends only on a relation between less complex verifications of A and of B.

and so when introducing a new term, if a verification of it can be defined only in terms of less complex verifications then its definition is not viciously circular. We need only to find the location of a sentence involving the new term on the hierarchy of complexity and to find its verification conditions we follow its dependencies down to the atomic cases.

Dummett must fear any form of holism for, on a holistic picture, the addition of new terms could cause a change in meaning in previously used terms. This would mean that a verification of a complex expression need not depend *only* on verifications of its parts, so the circularity cannot be removed by an appeal to a hierarchy of complexity of sentences (formulae).[19] The requirement Dummett puts on a theory of meaning is that the addition of a term of the language to any part of the language not containing that term

[18]This is because everything depends on use in a Dummettian theory of meaning, what counts as a verification also, in part, depends on how the terms of the language are used. Dummett, being an antirealist, does not wish to appeal to objective truth or validity conditions.

[19]Such considerations also lead Dummett to make his 'fundamental assumption' in [**?**]: A can be justified only by an application of the introduction rule for its main connective. Since I assume an innate concept of an objective truth, I am free from such worries about verification and do not require the fundamental assumption.

must constitute a *conservative* extension. The meaning of the new term must depend on its relation to the old terms without altering their meanings.[20]

The response to the sceptical worries for Dummett is quite simple. An implicit definition is legitimate if it conservatively extends an inference system. Put in terms of inferential roles, the genuine inferential roles are the ones that yield conservative extensions when a connective is added to fulfil them. Conservativeness can be characterised partly in terms of a proof theoretic property, what Dummett calls harmony.

I have shall give little objection to Dummett's account in this thesis, except perhaps its apparent reliance on inferential roles (which I have already discussed in a previous chapter). I do not wish to get involved in a discussion of what exactly Dummett's philosophy of language is, so I shall not engage Dummett much further. In later chapters I suggest methods by which we may obtain harmonious rules for classical logic, in particular a logic that contains the law of excluded middle. This suggests that one may maintain Dummett's verificationism and still be a realist in accepting bivalence (or at least the law of excluded middle), and so Dummett's philosophy need not be taken (on Dummett's on grounds) as an antirealist one. But I shall not push this point much in the thesis, my main concern now is to see how much logic we can get maintaining the demand for harmony.

I think it is interesting to note that considerations starting from opposing philosophical positions (realism and antirealism) both have a similar condition on the legitimacy of an implicit definitions. Dummett explicitly uses harmony (and is lead there by his verificationalism), and Peacocke's realism leads him to a demand for 'limiting principles' which is close to a demand for harmony (the fact that is is not a demand for harmony causes Peacocke his problems).

5.3 Concluding remarks

The demand for harmony has an elegance that should not be overlooked. Harmony is a proof theoretic property and does not depend on a semantics. Therefore the problems of assuming a pre-existing semantics that Peacocke suffered are avoided. Furthermore the demand for conservativeness guarantees us that the addition of the new term will not allow us to infer any new theorems of the old language. That is, when we add the new term there is no possibility of allowing us to infer something new that might be subject to empirical refutation. Of course, it might allow us to infer some old belief

[20]If it did then the definition of the new term would be viciously circular.

(that we could infer before making the definition) more easily,[21] but since the new term (and its definition) was not necessary for this inference, any problems with this old belief are not the fault of the definition.

Suppose t is a new term that has been introduced conservatively, and that the sentences P and C do not contain t. Since the addition of t (and its rules) is conservative it follows that if there is a truth preserving deduction of C from P using the term t (in the body of the deduction) then there is a truth preserving deduction of C from P without use of the term t at all. This means that we are free to *stipulate* that the rules of t preserve truth, a case of this stipulation leading us to a false belief that we did not already, at least tacitly, posses cannot arise.

What does this stipulation amount to? Here is a realist picture due to David Lewis.[22] A language has meaning and truth conditions in terms of sets of possible worlds (the famous picture), our conventions dictate which language we come to possess. So I claim that on Lewis' picture we may regard the stipulation of the rules of new term serves as part of the specification of which language (or languages) we possess (or in Lewis' terminology, which languages are those in which we place our trust). Since t is a conservative extension there really is a language that contains it together with all our old terms.[23]

[21] This is my answer to the question of what use deduction is, if there is a sense in which it provides us with no new information.

[22] See *Languages and Language*, collected papers Vol I.

[23] The problem of legitimate implicit definitions may be reformulated, on Lewis' picture, as the problem on which stipulations are guaranteed to have languages that satisfy them. Lewis' picture is good enough for me, although I shall say nothing about the nature of propositions.

Chapter 6

A first postulate for the rules of logic

Now that we have discussed some objections and replies to an implicit definition theory of analyticity, it is necessary to suggest what the definitions might be. The usual natural deduction rules for classical logic are a good start.

6.1 The rules

Here are some famous natural deduction rules:

$$\dfrac{A \wedge B}{A}\ \wedge E \qquad \dfrac{A \wedge B}{B}\ \wedge E \qquad \dfrac{A \quad B}{A \wedge B}\ \wedge I$$

$$\dfrac{A \vee B \quad \begin{array}{c}\cancel{A}^n \\ \vdots \\ C\end{array} \quad \begin{array}{c}\cancel{B}^n \\ \vdots \\ C\end{array}}{C}\ \vee E(n) \qquad \dfrac{A}{A \vee B}\ \vee I \qquad \dfrac{B}{A \vee B}\ \vee I$$

$$\dfrac{A \quad A \to B}{B}\ \supset E \qquad \dfrac{\begin{array}{c}\cancel{A}^n \\ \vdots \\ B\end{array}}{A \supset B}\ \supset I(n)$$

$$\dfrac{\bot}{B}\ \bot E$$

$$\frac{\sim\sim A}{A}\ DNE$$

where $\sim A$ is $A \supset \bot$.

If we leave aside the rule DNE then the system as a whole normalises, that is, all the connectives are eliminable in the sense above. In other words, the introduction an subsequent elimination of a formula in any deduction may be removed (or reduced). Any part of the deduction like this

$$\cfrac{\cfrac{\begin{matrix}\vdots & \\ A & B\end{matrix}}{A \wedge B}\ \wedge I}{A}\ \wedge E$$

may be replaced simply by the Prawitz tree at the beginning:

$$\begin{matrix}\vdots\\ A\end{matrix}$$

Any part of a deduction like this:

$$\cfrac{\cfrac{\begin{matrix}\vdots\\ A\end{matrix}}{A \vee B} \quad \begin{matrix}\cancel{A}^m\\ \vdots\\ C\end{matrix} \quad \begin{matrix}\cancel{B}^m\\ \vdots\\ C\end{matrix}}{C}\ \vee E(m)$$

may be reduced by appending the Prawitz tree $\begin{matrix}A\\ \vdots\\ C\end{matrix}$ (obtained by removing the crossing-out and superscript of $\ulcorner \cancel{A}^m \urcorner$) to copies of the initial Prawitz tree $\begin{matrix}\vdots\\ A\end{matrix}$, the whole thing is replaced by:

$$\begin{matrix}\vdots\\ A\\ \vdots\\ C\end{matrix}$$

A part like this:

$$\cfrac{A \quad \cfrac{\begin{matrix}\cancel{A}^n\\ \vdots\\ B\end{matrix}}{A \supset B}\ \supset I(n)}{B}\ \supset E$$

may be replaced by appending the Prawitz tree $\begin{matrix} A \\ \vdots \\ B \end{matrix}$ (obtained similarly by removing crossings-out and superscripts) to copies of the Prawitz tree $\begin{matrix} \vdots \\ A \end{matrix}$:

$$\begin{matrix} \vdots \\ A \\ \vdots \\ B \end{matrix}$$

there is no case to consider for $\perp$ as it has no introduction rule.

Now consider any formula X that is the conclusion of an introduction rule and is the conclusion of a minor premise of $\vee E$, and is not the conclusion of the deduction:

$$\dfrac{\dfrac{A \vee B \quad \begin{matrix} \cancel{A}^n \\ \vdots \\ X \end{matrix} \quad \begin{matrix} \cancel{B}^n \\ \vdots \\ X \end{matrix}}{X}\,\vee E(n)}{D}\,?$$

X is the conclusion of $\vee E$ and is then used in some other rule. We replace such a section of a deduction by:

$$\dfrac{A \vee B \quad \begin{matrix} \cancel{A} \\ \vdots \\ \dfrac{X}{D}\,? \end{matrix} \quad \begin{matrix} \cancel{B} \\ \vdots \\ \dfrac{X}{D}\,? \end{matrix}}{D}\,\vee E$$

by repeatedly making these replacements we ensure that no conclusion of an introduction rule which is not the conclusion of the deduction is also the conclusion of the minor premise of an application of $\vee E$.

Now let X be a formula in a deduction that is (i) the conclusion of an introduction rule, (ii) is not the conclusion of the deduction, and (iii) has at least as high a degree (see 1.4.2) as any other formula satisfying (i-ii). This formula X is must be eliminated next step as any introduction rule increases degree (and we have shown above that it is not used in the deductive premise of a $\vee E$). Applying the simple reduction cases above this introduction and subsequent elimination may be removed. By induction on the degree of X it follows that a deduction may be normalised (all introductions and subsequent eliminations of the same formula may be removed).

The problem here is that we proved normalisation for a system without DNE, which is a classically valid rule and is not derivable from the other rules. In the system postulated above there are introductions and subsequent eliminations of $\supset$ that we cannot remove:

$$\dfrac{\dfrac{\begin{array}{c}\cancel{\sim A}^{n}\\ \vdots\\ \bot\end{array}}{\sim\sim A}\supset I(n)}{A}\,DNE$$

and we cannot use the inference from $\sim A$ to $\perp$ to obtain a direct inference of A. Consider this example:

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\cancel{A}^{2}}{A \vee \sim A}\vee I \quad \cancel{\sim(A \vee \sim A)}^{1}}{\bot}\supset E}{\sim A}\supset I(2)}{A \vee \sim A}\vee I \quad \cancel{\sim(A \vee \sim A)}^{1}}{\dfrac{\bot}{\sim\sim(A \vee \sim A)}\supset I(1)}}{A \vee \sim A}\,DNE$$

and we cannot reduce this to a direct deduction of $A \vee \sim A$ without the introduction and subsequent elimination if $\sim\sim(A \vee \sim A)$.

Scrapping the rule DNE allows normalisation but deprives us of classical logic. I will turn to my chosen method of obtaining normalisation for classical logic in chapter 8.

6.2 A summary of the story so far

I say that we may define a connective by making an implicit definition. I have not stated by what means the definition is made, though I suggest a theory of it in chapter 14.

The condition on a legitimate implicit definition of a connective or a system of connectives is that of harmony.

The requirement of harmony extends only to logical connectives that are defined entirely by their inference rules. Not all connectives have this property, for example temporal connectives obtain their meaning, in part, by our perception of of time (and perhaps also change). Thus the content of a temporal or special operator is not entirely a logical matter. Consequently we should not be surprised if the logic of temporal operators is not harmonious (or conservative over our basic logic).

We should be surprised (if I am correct about implicit definitions) if our basic system of deduction and entailment is not harmonious. At the moment, this claim derives from an intuition on my part, I see no other general condition of legitimacy on implicit definitions and I believe the implicit definition account of analyticity is correct.

Chapter 7

Normalisation for classical logic: other attempts

In response to Dummett's demand (and the seemingly general need that any implicit definition theory has) for normalisation of deductions (a.k.a. harmony) Rumfitt presents an enhancement to the standard Prawitz natural deduction structure so that classical logic normalises. A good, if not the only, reason to demand normalisation is to ensure consistency, however, I show that with Rumfitt's enhancements normalisation does not entail consistency.

7.1 Why normalisation

One response to problems of the justification of deduction is that the logical connectives are implicitly defined by their introduction and elimination rules. The deductive inferences are valid because logical connectives are implicitly defined so that they are valid. However, not all implicit definitions are legitimate. Prior's connective *tonk* is implicitly defined so that

$$\frac{A}{A\ tonk\ B}\ tonk I \qquad \frac{A\ tonk\ B}{B}\ tonk E$$

are valid. But in a system with these inference rules all propositions become equivalent.

A natural response notes that since B is a consequence of $A\ tonk\ B$ we should expect B to be necessary for introducing it. Think of $A\ tonk\ B$ as a package containing some data, according to the elimination rule it contains B. We should expect then that B should have been required for the production of the package $A\ tonk B$. But it was not. Compare this with conjunction

$$\frac{A \quad B}{A \wedge B} \qquad \frac{A \wedge B}{A} \qquad \frac{A \wedge B}{B}$$

the consequences of $A \wedge B$ are part of (in this case exactly) what is required to introduce it. Similarly with implication:

$$\frac{\begin{array}{c}\cancel{A}\\ \vdots \\ B\end{array}}{A \rightarrow B} \qquad \frac{A \rightarrow B \quad A}{B}$$

the consequence of $A \rightarrow B$ is that we can infer B from A, and this is precisely what we need to introduce it.

A suggested requirement on a system of implicit definitions (of logical connectives) is then that deductions in it normalise. That is, no connective is required to be introduced and then later be eliminated. If this is the case then we can be sure that the elimination rules for the connectives allow us to infer only what was required by their introduction rules. If the system normalises in this way then it is consistent. To see this, suppose $\Gamma \vdash A$ where A is atomic and Γ contains only atomic propositions. If deductions normalise then there is a deduction of A from Γ which involves no introduction and subsequent elimination of any connective. Since the deduction that $\Gamma \vdash A$ begins with atomic premises and ends with an atomic conclusion there is a deduction of A from Γ involving no logical connectives at all.[1] But in Prawitz natural deduction systems (assumed to be the framework for deductions) such a deduction can occur only if $A \in \Gamma$.

7.2 Principle of indirect proof

A well known way of retaining normalisation of classical logic is to include the Principle of indirect proof PIP:

$$\frac{\begin{array}{c}\neg A\\ \vdots \\ \bot\end{array}}{A}$$

The formal theory behind this rule may be seen in Prawitz's book [**?**] and Stalmark's paper [**?**].

[1]Since any connective introduced must then be eliminated (so that the conclusion is atomic) there is a deduction in which no connective is introduced.

In [**?**], Stalmark proves Strong Normalisation for the formulation of classical logic that uses a more general version Prawitz' classical absurdity rule (i.e. the principle of indirect proof, PIP).

Dummett considers this rule in [**?**, p296-300] and concludes that it cannot be a rule that defines negation (or any other connective) at least within his theory of how logical connective obtain their meaning. The basic problem for Dummett is that with PIP, the logical connectives can no longer be said to be defined by only their own inference rules. For example we can use PIP, a rule for negation and absurdity only, to deduce the Peirce's law $((A \rightarrow B) \rightarrow A) \rightarrow A$. Indeed, in the systems of Stalmark and Prawitz we *must* use PIP to deduce $((A \rightarrow B) \rightarrow A) \rightarrow A$. But in Dummett's view, if $((A \rightarrow B) \rightarrow A) \rightarrow A$ is a theorem of our basic logic, it must be deducible from only the structural rules of deduction and the rules for $\rightarrow$ (the only connective that appears in the formula in question).

The theory of implicit definition that I have presented so far is not as rigid as Dummett's. So far I have demanded only normalisation of the system of logical connectives as a whole see 3.3. So, given only what I have argued up to this point, the only objection I have is that the rule PIP is inelegant (for the same reason that Dummett rejects it, it is an inelegancy that a negation rule must be used to deduce an implication only formula).

However, in chapter 14 I propose a more detailed theory of what implicit definitions are and how they are made. I argue that an implicit definition must consist only of introduction and elimination rules utilising only the defined connective which itself appears only in the premises of the elimination rules and the conclusions of the introduction rules. My arguments for this are along the same lines as the arguments given by Hacking in [**?**].

My problem with PIP is that it is neither an introduction nor an elimination rule (see section 1.3.1 of this book). The significance of this to my account of implicit definition is that the rule cannot be then said to be defining negation in terms of some purely structural property of deductions (see chapter 14).

It is true that Prawitz' and Stalmarck's rule can be viewed as an elimination rule for absurdity. Such an interpretation of the rule once again does not fit with my thoughts on implicit definitions. The worry is that absurdity, which features in the negation introduction rule, becomes itself defined in terms of negation. This makes it hard to argue that each logical connective is (individually) implicit in the structure of deduction (and the idea that the logical connectives are already there, waiting to be defined, is important to my account of analyticity, again see chapter 14).

Since chapter 14 is not till much later in this thesis, I cannot yet object to a theory of analyticity that follows mine up to this point, but makes use of PIP. The most I can object so far is that there is an inelegancy in PIP. If it is a negation rule then it is inelegant because neither an introduction nor an elimination rule for negation, but if it is not a rule for negation then it is inelegant because negation appears in the premise. In chapter 8 I present the Restart rule which is fully in the spirit of PIP except it is a purely structural rule (it contains no connectives at all, not even $\perp$). The restart rule is more elegant, but more importantly, being a structural rule, it fits well with the theory of analyticity I give in chapter 14.[2]

Indeed, the similarity of the Restart rule (which I shall favour) to PIP shows that PIP is along the right lines.

7.3 Enhancements of Prawitz natural deduction

The usual deduction rules for classical logic do not normalise. In order to deduce $A \vee {\sim}A$ from empty assumptions we must introduce and eliminate a negation by deducing that ${\sim}{\sim}(A \vee {\sim}A)$ and then finishing with double negation elimination.

We can obtain normalisation for classical logic by enhancing the basic structure of inferences. Read in [?] does so by giving natural deduction multiple conclusions. Rumfitt in [?] does so by adding signing the nodes in the deduction tree with $+$ or $-$ and adding special *structural reductio* rules.[3]

---

[2] And, also because it is structural, the Restart rule will fit better with the theories of Dummett and Hacking.

[3] We may play the structural-rule-reductio game another way, here with a more apparently structural property. Allow formulae to be expressed as usual and also by writing things upside down. We then add the following structural rules:

$$\dfrac{A \quad \forall}{B}\ \textit{Abstraction} \qquad \begin{array}{c} \cancel{\forall}^{n} \\ \vdots \\ \dfrac{A}{A} \end{array} \textit{Spinning}(n)$$

If we disallow the syntactic variables from varying over formulae the other way up (i.e interpreting the configuration of the inference rules literally) then every formula will be either completely the right way up or the wrong way up. Whatever we choose to disallow:

$$\dfrac{\dfrac{\dfrac{\dfrac{\cancel{\forall{\sim}\wedge\forall}^{2} \quad \dfrac{\cancel{A}^{1}}{A \vee {\sim}A}\vee I}{\perp}\textit{Abstraction}}{{\sim}A}\to I(1)}{A \vee {\sim}A}\vee I}{A \vee {\sim}A}\textit{Spinning}(2)$$

$$\dfrac{\begin{array}{cc}\cancel{-A}^{n} & \cancel{-A}^{n}\\ \vdots & \vdots\\ +B & -B\end{array}}{+A}\,SR(n) \qquad \dfrac{\begin{array}{cc}\cancel{+A}^{n} & \cancel{+A}^{n}\\ \vdots & \vdots\\ +B & -B\end{array}}{-A}\,SR(n)$$

In Rumfitt's system the structural properties of deductions have been enhanced so that some non-premise circular deductions can be made by structural rules alone.[4] For example the inference

$$\dfrac{+A \quad -A}{+B}$$

by empty discharging $-B$ (with an application of the structural reductio rule) shows that anything follows from any assumptions that include $\{+A, -A\}$ (this is a form of *ex falso*).

The structural enhancements are not without their problems. For Read a major difficulty is that people seem to reason in terms of single conclusions rather than multiple conclusions.[5] For Rumfitt's system the difficulty is more serious; normalisation no longer entails consistency. In which case the demand for normalisation is not a sufficient condition on legitimacy of implicit definitions and we have advanced nowhere. At least, anyone seeking a proof theoretic justification of the logical constants cannot use Rumfitt's system as it stands.

Consider the 0-place connective $\bullet$, its rules are:

$$\dfrac{-\bullet}{+\bullet}\,\bullet 1 \qquad \dfrac{+\bullet}{-\bullet}\,\bullet 2$$

since $\bullet$ has no introduction rule it can never be introduced and then eliminated. Furthermore the rules have a symmetry to them which Rumfitt

is a normalised deduction.

[4]In a multiple conclusion logic the inference $A \vdash A, B$ is premise circular as one of the multiple conclusion is contained in the premises.

[5]For example the required introduction rule for implication

$$\dfrac{\begin{array}{c}\cancel{A}^{n}\\ \vdots\\ B, C\end{array}}{A \supset B, C}\,\supset I(n)$$

is not something I believe I have ever used. The multiple conclusion should be read like a disjunction so that the introduction rule $\supset I$ adds that $A \supset (B \vee C)$ entails $(A \supset B) \vee C$.

regards as a mark of a well defined connective: each rule is the mirror image of the other. However:

$$\dfrac{\dfrac{\dfrac{\cancel{-\bullet}^1}{+\bullet}\bullet 1 \quad \cancel{-\bullet}^1}{+\bullet}SR(1) \qquad \dfrac{\dfrac{\dfrac{\cancel{-\bullet}^2}{+\bullet}\bullet 1 \quad \cancel{-\bullet}^2}{+\bullet}SR(2)}{-\bullet}\bullet 2}{+C}SR$$

is a deduction of $+C$ from no premises for any C.[6] So Rumfitt cannot argue that normalisation (or a pleasing symmetry to the rules) guarantees a legitimate implicit definition.

Demanding that the connective appear on only one side of its rules provides no relief.

$$\dfrac{+A}{-\bullet}\bullet I \qquad \dfrac{-A}{+\bullet}\bullet I$$

$$\dfrac{+\bullet \quad \pm C}{\pm C}\bullet E \qquad \dfrac{-\bullet \quad \pm C}{\pm C}\bullet E$$

Clearly introduction and subsequent elimination of $\bullet$ is unnecessary. Nevertheless:

$$\dfrac{\dfrac{\dfrac{\cancel{+A}^1}{-\bullet}\bullet I}{+\bullet}\bullet I \quad \dfrac{\cancel{+A}^1}{-\bullet}\bullet I}{-A}SR(1)$$

is a deduction of $-A$ for any A, and similarly we get a deduction of $+A$ for any A. The situation is worse still, consider these rules:

$$\dfrac{+A \quad -A}{-\bullet}\bullet I \qquad \dfrac{+A \quad -A}{+\bullet}\bullet I$$

$$\dfrac{+\bullet}{\pm A}\bullet E \qquad \dfrac{-\bullet}{\pm A}\bullet E$$

notice that the introduction rules for $\bullet$ are *valid* by an application of SR in Rumfitt's system independently of whether we have defined any rules for $\bullet$ (as anything follows from $\{+A, -A\}$). The elimination rules clearly add no more than the introduction rules yet:

$$\dfrac{\dfrac{\dfrac{\cancel{+\bullet}^1}{+A}\bullet E \quad \dfrac{\cancel{+\bullet}^1}{-A}\bullet E}{-\bullet}SR(1)}{\pm A}\bullet E$$

[6]The final step involves the empty discharge of $-C$ by the rule SR. We could also empty discharge $+C$ to obtain a deduction of $-C$.

is a deduction of $+A$ and $-A$ for any A. It is left for Rumfitt to explain why rules that are valid independently of any stipulations about meaning are illegitimate for making an implicit definition.[7]

A solution to this problem (for Rumfitt's system) lies in further restricting what rules can be used to define a connective. The definitional introduction rules should have only a positive conclusion (labelled by $+$), and the definitional elimination rules should have only positive minor premises. So for example the definitional rules for negation are:

$$\frac{-A}{+\sim A} \qquad \frac{+\sim A \quad +A}{\bot}$$

With this restriction normalisation does entail consistency (a proof of this is not necessary would just take up space here). The problem with this restriction is that it seems *ad hoc*, there seems no independent motivation for such a restriction. After all, it seems to be the point of Rumfitt's system that a connective is defined by its falsity as well as its truth conditions (hence $-$ and $+$ in the structure of the deductions). A second problem may arise as the rules get more complicated and it becomes less clear exactly how the restriction is to apply (e.g. when it is not clear what the major premises are, or where there are rules that are neither introduction nor elimination rules).

7.4 Conclusion

If we wish to make much use of implicit definitions then it is paramount that we can rule out the many examples of implicit definitions leading to disaster (e.g. tonk). A demand for harmony will ensure consistency and provide us with non-semantic conditions on when an implicit definition is legitimate. It is important to note that harmony is a property of the structure of an implicit definition itself rather than of what is defined, i.e. harmony is not a semantic property, so we are not assuming a semantics when we implicitly define our logical connectives (and from them our semantics). We then have the basics of an extremely elegant response to the problem of the justification

[7] In the usual system (without $+, -$) we can easily say for what is wrong, the analogues of the rules are:

$$\frac{A \quad \sim A}{\sim\bullet} \qquad \frac{A \quad \sim A}{\bullet}$$

$$\frac{\bullet}{A} \quad \frac{\bullet}{\sim A} \qquad \frac{\sim\bullet}{A} \quad \frac{\sim\bullet}{\sim A}$$

and we have added an extra elimination and introduction rules for $\sim$.

of deduction: the connectives are implicitly defined by consistent implicit definitions, a definition is consistent when it yields normalising deductions. As far as I can see there is no reason to require normalisation other than to be able to run such a line.

Since in Rumfitt's system normalisation does not entail consistency we cannot run such a line, and there is no reason to demand normalisation anymore. There is also no response to the problems of tonk and $\bullet$ and we have no criterion for ruling out some implicit definitions as illegitimate (aside from begging the whole question and assuming a semantics with which to determine which rules are consistent).

7.4.1 A comment on $\bullet$

The connective $\bullet$ comes from an ingenious argument in [**?**], that a form of symmetry between introduction and elimination rules does not guarantee normalisation or consistency. Stephen Read presents the connective $\bullet$.

$$\frac{\sim\bullet}{\bullet}\ \bullet I \qquad \frac{\bullet \quad \begin{array}{c}\cancel{\sim\bullet}^{n}\\ \vdots \\ C\end{array}}{C}\ \bullet E(n)$$

These rules yield an inconsistency, together with the negation rules. Notice that $\bullet \vdash \sim\bullet$ with an application of the elimination rule $\bullet E$ to $\{\bullet, \sim\bullet\}$ and then discharging $\sim\bullet$:

$$\frac{\bullet \quad \cancel{\sim\bullet}^{1}}{\sim\bullet}\ \bullet E(1)$$

so

$$\dfrac{\dfrac{\dfrac{\begin{array}{c}\cancel{\bullet}^{1}\\ \vdots \\ \sim\bullet\end{array} \quad \cancel{\bullet}^{1}}{\bot}\ \sim E}{\sim\bullet}\ \sim I(1) \qquad \dfrac{\dfrac{\dfrac{\begin{array}{c}\cancel{\bullet}^{2}\\ \vdots \\ \sim\bullet\end{array} \quad \cancel{\bullet}^{2}}{\bot}\ \sim E}{\sim\bullet}\ \sim I(2)}{\bullet}\ \bullet I}{\bot}\ \sim E$$

A similar result may be found in the following naive set theoretic principles, which more obviously have the pleasing symmetry.

$$\frac{A(a)}{a \in \{x : A(a/x)\}}\ SI \qquad \frac{a \in \{x : A(a/x)\}}{A(a)}\ SE$$

Russell's paradox now follows similarly to Read's example with $\bullet$.

$$\dfrac{\dfrac{\dfrac{\dfrac{\cancel{\{x : x \notin x\} \in \{x : x \notin x\}}^1}{\{x : x \notin x\} \notin \{x : x \notin x\}}\,SE \quad \cancel{\{x : x \notin x\} \in \{x : x \notin x\}}^1}{\bot}\,{\sim}E}{{\sim}(\{x : x \notin x\} \in \{x : x \notin x\})}\,{\sim}I(1) \qquad \dfrac{\dfrac{\dfrac{\dfrac{\cancel{\{x : x \notin x\} \in \{x : x \notin x\}}^2}{\{x : x \notin x\} \notin \{x : x \notin x\}}\,SE \quad \cancel{\{x : x \notin x\} \in \{x : x \notin x\}}^2}{\bot}\,{\sim}E}{{\sim}(\{x : x \notin x\} \in \{x : x \notin x\})}\,{\sim}I(2)}{\{x : x \notin x\} \in \{x : x \notin x\}}\,SI}{\bot}\,{\sim}E$$

Notice that in both these examples $\sim$ is introduced and then eliminated in the final step. This happens because the rules SI and $\bullet I$ give an alternative way of eliminating $\sim$. So after deducing e.g. ${\sim}\bullet$ we can eliminate the negation in an unusual way to obtain $\bullet$ and then reuse ${\sim}\bullet$ to obtain $\bot$ (eliminating $\sim$ in the usual way).

So, in fact, although the rules officially labelled the introduction and elimination rules for negation have a symmetry to them, the complete set of rules governing it does not. If we are liberal with our understanding of harmony then these are not counterexamples to harmony as a path to consistency. But notice that we are moving away from a harmony understood as a property of a single connective to a global property about the entire logic (a property of *all* the introduction and elimination rules together in the whose system).

Chapter 8

Harmony for classical logic: the restart rule

See [?] for a formulation of this rule in goal directed reasoning.

8.1 The rule

Here is the *restart* rule for natural deduction.

$$\frac{A}{B}\ \textit{restart}$$

Provided that A reoccurs below in the Prawitz tree.

That is, the side condition on a rule application of restart is met if there is another occurrence of A in the Prawitz tree below that rule application.

The restart rule is unusual in that any application of it depends on what is below it in a deduction rather than what is above it.

If we add the restart rule to the standard deduction system for intuitionistic logic we obtain a classical deduction system. Here are two examples of the restart rule at work.

$$\cfrac{\cfrac{\cfrac{\cfrac{\cancel{A}^1}{B}\ \textit{restart}}{A \supset B}\ {\supset}I(1) \qquad \cancel{(A \supset B) \supset A}^2}{A}\ {\supset}E}{((A \supset B) \supset A) \supset A}\ {\supset}I(2) \qquad\qquad \cfrac{\cfrac{\cfrac{\cfrac{\cancel{A}^1}{A \vee {\sim}A}\ {\vee}I}{\bot}\ \textit{restart}}{{\sim}A}\ {\sim}I(1)}{A \vee {\sim}A}\ {\vee}I$$

8.1.1 Some important definitions

If the restart rule is applied

$$\frac{A}{B}\ \textit{restart}$$

then lower down in the deduction A must be deduced again.

- If A is the premise of an application of restart, and B occurs below A, but no other occurrence of A occurs between B and A, say that the application of restart is *incomplete at* B. In other words, if B occurs above all occurrences of A that satisfy the side-condition on an application of restart above B, then that application of restart is incomplete at B.

- Say any formula occurring below an incomplete application of restart depends on an incomplete application of restart.

- An incomplete application of restart is an *weak rule application*. That is, if a certain application of restart is incomplete at B then it is a weak rule application at B.[1]

- Incomplete applications of restart are *assumptive rule applications*. That is, if a certain application of restart is weak (or incomplete) at B, then it is also *assumptive* at B (i.e. it is an assumptive rule application at B).[2]

- All the other rule applications I have defined so far are not assumptive (so complete applications of restart are not assumptive). [3]

- If a term t occurs in the premise of an application of restart we may say simply that it occurs in that application of restart. If the term is

[1] An application of restart is a *weak* rule application as long as it is incomplete.

[2] Intuitively, an assumptive rule application is a rule application that involves a hidden assumption that may be discharged (remaining hidden) later in the deduction. Restart is the only inference rule I shall define here that can have assumptive applications (though I shall define in 11.3 other rules any application of which are always weak).

[3] Assumptive rule applications are rule applications that involve hidden, or structural assumptions. A structural assumption is one where the assumption is a particular formula, rather it is an assumption of (the validity of) an inference pattern. In the case of restart the inference pattern is effectively this:

$$\frac{A}{\perp}$$

which is generalised by replacing $\perp$ by B.

a variable that is free in the premise, then we may say that it is free in the application of restart.[4]

- If an atomic predicate X occurs in the premise of an application of restart we may say simply that it occurs in that application of restart.

8.2 How to interpret it

In classical reasoning, the inference rule PIP is sound.

$$\dfrac{\begin{matrix}\cancel{\sim A}^{n}\\ \vdots \\ A\end{matrix}}{A}\; PIP(n)$$

We may discharge the assumption $\sim A$ if we can deduce A from it. With the assumption that $\sim A$ we may deduce B from a line in the proof containing A (as a conclusion). The restart rule may be seen as making the implicit assumption that $\sim A$. This assumption must then be discharged, for otherwise the inference would not be sound as the assumption is not explicitly made. The assumption may be discharged by concluding that A at some later point. Thus the rule is sound. We can make this reasoning more formal and show than any application of restart may be replaced in a deduction with an application of PIP. Thus we can think of restart as being another way of formulating PIP. But I think this is not the best interpretation of the restart rule for it requires that we interpret it in terms of another connective, $\sim$.

I think a better interpretation not that the restart rule makes an implicit assumption that $\neg A$, but that it makes an implicit assumption of a structural form:

$$\frac{A}{B}$$

or less generally

$$\frac{A}{\bot}$$

the rule then allows this assumption to be discharged when we deduce A again. On this interpretation the restart rule may still be seen as a form of indirect proof (PIP) without interpreting it in terms of $\sim$. That is, on this interpretation the restart rule may be seen as a fully structural rule.

[4]The structural assumption in the application of restart involves A, so we must regard the free variables of A as free in the structural assumption until it is discharged.

It is easy to see, therefore, that the restart rule is sound for classical semantics, for we may replace any instance of restart:

$$\frac{A}{B}\ \textit{restart}$$

with this

$$\dfrac{\dfrac{A \quad \cancel{\sim A}^n}{\bot}\sim E}{B}\ \bot E$$

and then replace the first re-occurrence of A below with

$$\frac{A}{A}\ PIP(n)$$

After all replacements have been made we obtain a deduction which make use of PIP and not restart. So anything deducible using restart is deducible using PIP, since PIP is sound for classical semantics, so is restart.

8.3 Normalisation

In the normal form argument there are new cases to consider Suppose $A \wedge B$ is introduced and then eliminated:

$$\dfrac{\dfrac{A \quad B}{A \wedge B}\wedge I}{A}\wedge E$$

then we may not be able to remove the introduction if $A \wedge B$ is introduced to complete a previous application of restart. For example the deduction might contain

$$\dfrac{\dfrac{\begin{matrix}\vdots \\ \dfrac{A \wedge B}{C}\ \textit{restart} \\ \vdots \\ A\end{matrix} \qquad \begin{matrix}\vdots \\ B\end{matrix}}{A \wedge B}\wedge I}{A}\wedge E$$

and if we reduce this to:

$$\begin{matrix}\vdots \\ \dfrac{A \wedge B}{C}\ \textit{restart} \\ \vdots \\ A\end{matrix}$$

then the inference as a whole may no longer be a deduction as the application of restart may remain incomplete. In order to remove the unnecessary conjunction introduction we must first push the application of restart down a step in the deduction:

$$\dfrac{\dfrac{\begin{array}{c}\vdots\\ \dfrac{\dfrac{A \wedge B}{A}\,\wedge E}{C}\,\mathit{restart}\\ \vdots\\ A\end{array} \qquad \begin{array}{c}\vdots\\ B\end{array}}{A \wedge B}\,\wedge I}{A}\,\wedge E$$

and now the application of restart is completed by A and not by $A \wedge B$, the introduction and subsequent elimination of $A \wedge B$ may then be removed as usual.

The elimination rules are of the form:

$$\dfrac{P \quad M_1 \quad M_2}{C}$$

where P is the major premise and each M_i are the minor premises (the M_i are either formulae or inferences). Suppose that P completes an earlier application of restart. Then part of the deduction looks like this:

$$\dfrac{\begin{array}{c}\dfrac{P}{X}\,\mathit{restart}\\ \vdots\\ P\end{array} \qquad \begin{array}{c}\vdots\\ M_1\end{array} \quad \begin{array}{c}\vdots\\ M_2\end{array}}{C}$$

we may then we need only change the deduction so that restart is applied to the conclusion of the elimination rule and not its premise:

$$\dfrac{\begin{array}{c}\dfrac{\dfrac{P \quad \begin{array}{c}\vdots\\ M_1\end{array} \quad \begin{array}{c}\vdots\\ M_2\end{array}}{C}}{X}\,\mathit{restart}\\ \vdots\\ P\end{array} \qquad \begin{array}{c}\vdots\\ M_1\end{array} \quad \begin{array}{c}\vdots\\ M_2\end{array}}{C}$$

and now the application of restart is completed by C and not by P. Not also the node that completes the restart rule is one deduction step closer to the conclusion of the whole deduction.

We may therefore ensure that any deduction may be reduced to a deduction where no formula both completes an application of restart and is the premise of an elimination rule. Then the deduction may be normalised by the usual methods, for if a connective is introduced and then eliminated we know that the introduction did not serve to complete an application of restart and may be removed without affecting the legitimacy of the deduction.

We must still argue that normalisation entails consistency as the restart rule has, in effect, a structural enhancement. To sketch the argument suppose that atomic A is a conclusion of only atomic premises Γ. Now consider a legitimate application of restart:

$$\begin{array}{c} \dfrac{P}{C}\ \textit{restart} \\ \vdots \\ P \end{array}$$

(the application is legitimate only if P is deduced later in the deduction) if none of the premises on which P depends are discharged in between C and P then the application is unnecessary and the deduction between C and P may be deleted. Thus of A is deducible from Γ using only applications of restart then $A \in \Gamma$. If the other connectives satisfy a normal form theorem then $\Gamma \vdash A$ only if the deduction can be made without use of any introduction or elimination rules at all (provided Γ and A are all atomic). But with only restart we can deduce only members of Γ from Γ.

8.4 Remarks

The restart rule might seem initially like clever trick with the syntax that merely trivialises the demand for harmony. The restart rule is by no means trivial, it is a way of introducing negative hypothesis to the structure of deduction without requiring negation. Thus we may temporarily assume that a proposition is false simply by applying restart to it. This hidden assumption is discharged when we complete the restart rule. Certainly we use indirect proof often in our reasoning, it is a difficult principle to avoid, if assuming the butler did not do it leads to absurdity then the butler did do it. Restart allows such reasoning, in particular it allows the negative assumption, to be structural rather than in terms of rules for the connectives. Thus indirect proof (PIP) may be understood not as essential to the meaning of negation but as essential to the structure of deduction. The restart rule is a form of indirect proof that does not require negation or any other connectives to be formulated.

If we are willing to accept that restart rule is part of actual reasoning (we need only accept that we sometimes when we apparently use PIP we are actually using the restart rule but cannot explain our reasoning easily to others without describing it as PIP), then we have a harmonious natural deduction system for classical logic (harmonious in both Dummett's senses, see page 50). This means that we may, after all, give a proof theoretic account of the meaning of the logical constants (in terms of implicit definitions) which yields classical logic (rather than the weaker intuitionistic logic).

Chapter 9

Sheffer Stroke

I shall now present another method of obtaining a normalisation theorem for classical logic. The inference rules are presented for a single conclusion Prawitz style natural deduction system. The connectives are defined holistically in terms of Sheffer stroke. The introduction rule for Sheffer stroke satisfies the subformula property (the premises are subfomrulae of the conclusion). However the elimination rule is such that some discharged assumptions may not be subformulae of the premises or the conclusion of the deduction as a whole.

A reason for demanding the subformula property of the introduction rule is so that it can function as a non-circular definition of the connective introduced. This is perhaps a well motivated demand, although not all circular definitions are vicious (for example an inductive definition). Dummett also requires that every formula in a deduction be a subformula of a premise or the conclusion ([**?**, p281]). Dummett's reason for demanding this is what he calls the *fundamental assumption*, which even Dummett accepts is a questionable principle ([**?**, ch12]). The assumption is that any justification for a sentence A must ultimately come from introducing the main connective of A by an introduction rule.[1] The assumption, if it is fundamental at all, is fundamental to Dummett's verificationist account of meaning and not to my account of analyticity, so far I have given an account of analyticity (in terms of implicit definitions) which makes no use of Dummett's fundamental assumption. Consequently I see no reason to adhere to Dummett's demand that every formula of a deduction be a subformula of a premise or

[1]This is highly problematic, as Dummett observes, for example we can know $A \vee B$ without having introduced it by means of a prior knowledge of which disjunct is true (as the introduction rule requires).

the conclusion.

9.1 The propositional case

In the following sections I show normalisation and completeness for a deduction system for Sheffer stroke. Note that the system I construct does not make any use of the restart rule. Since I shall favour the restart rule as the preferred method of obtaining normalisation for classical logic, this is the only chapter of this thesis that does not assume the restart rule.

9.1.1 Sheffer stroke: semantics and completeness

Consider the following connective, called Sheffer stroke (denoted by $|$). Here are two natural deduction systems for it, the consequence relations of which are denoted by $\vdash_I$ and $\vdash_C$:

The system *Int* ($\vdash_I$)

$$\dfrac{\begin{array}{c}\cancel{A}^n \quad \cancel{B}^n\\ \vdots \\ \bot\end{array}}{A|B}\ II(n) \qquad \dfrac{A|B \quad A \quad B}{C}\ IE$$

$$\dfrac{\bot}{B}\ \bot E$$

The system *Clas* ($\vdash_C$)

$$\dfrac{\begin{array}{c}\cancel{A}^n \quad \cancel{B}^n\\ \vdots \\ \bot\end{array}}{A|B}\ CI(n) \qquad \dfrac{A|B \quad \begin{array}{c}\cancel{C|C}^n\\ \vdots\\ A\end{array} \quad \begin{array}{c}\cancel{C|C}^n\\ \vdots\\ B\end{array}}{C}\ CE(n)$$

$$\dfrac{\bot}{B}\ \bot E$$

We add rules for conjunction to both *Clas* and *Int*:

$$\dfrac{A \quad B}{A \wedge B}\ \wedge I \qquad \dfrac{A \wedge B}{A}\ \wedge E \qquad \dfrac{A \wedge B}{B}\ \wedge E$$

conjunction is not necessary for the proof theory, but it is hard to doubt that we use conjunction defined directly by these rules rather than in terms of Sheffer stroke.

It is important to be aware of empty discharging. For example, we may deduce $(A|A)|B$ from A:

$$\dfrac{\dfrac{A \quad \cancel{A|A}^1}{\perp}\,IE}{(A|A)|B}\,II(1)$$

which involves the empty discharge of B.

Theorem 9.1.1 If $\Gamma \vdash_I A$ then $\Gamma \vdash_C A$

Proof: any use of the I-elimination rule (IE)

$$\dfrac{\begin{matrix}\vdots \\ B|C\end{matrix} \quad \begin{matrix}\vdots \\ B\end{matrix} \quad \begin{matrix}\vdots \\ C\end{matrix}}{D}\,IE$$

may be relabelled as CE to yield an application of CE where $D|D$ is empty-discharged. ∎

The converse does not hold. *Int* is sound for the interpretation of $A|B$ as $\neg(A \wedge B)$ (where $\neg$ is intuitionistic), whereas *Clas* is sound for $A|B$ interpreted as $\sim(A \wedge B)$ (where $\sim$ is classical).

Semantics

The interpretation of (classical) Sheffer stroke is given by this truth table.

A	$\|$	B
T	F	T
T	T	F
F	T	T
F	T	F

Also called *alternative denial* Sheffer stroke is true when A is false or B is false, where classical truth valuations apply (i.e. NAND). Note that $A|A$ is true when A is false. A reading of $|$ into natural language is perhaps '. . . is (accidentally) incompatible with. . . ' (accidentally as opposed to necessarily). Classical models are defined in terms of truth valuations satisfying the above truth table.

Completness

First we will show that interpreting truth as membership in a maximal consistent set of *Int*, call it M,[2] yields a classical model, i.e. that

(†) $A|B \in M$ iff $A \notin M$ or $B \notin M$.

First see that $\{A, B, A|B\}$ is a contradictory set by a single application of IE:

$$\frac{A|B \quad A \quad B}{C}\ IE$$

this proves the left-right direction of †. Taking $A = B$, $\{A, A|A\} \vdash_I \perp$ and also $\{B, A, A|A\} \vdash_I \perp$, so $A|A \vdash_I A|B$ and similarly $B|B \vdash_I A|B$:

$$\frac{\dfrac{A|A \quad \cancel{A}^1}{\perp}\ IE}{A|B}\ II(1) \qquad \frac{\dfrac{B|B \quad \cancel{B}^1}{\perp}\ IE}{A|B}\ II(1)$$

Assume that

(‡) $A \in M$ or $A|A \in M$, for any A

then if $A \notin M$ or $B \notin M$ then either $A|A \in M$ or $B|B \in M$ and so, by the deductions above, $A|B \in M$. This shows the right-left direction of †.

‡ follows, by maximal consistency of M, from the fact that

if $\Gamma, A|A \vdash_I \perp$ and $\Gamma, A \vdash_I \perp$ then $\Gamma \vdash_I \perp$

for if $\Gamma, A \vdash_I \perp$ then $\Gamma \vdash_I A|A$ by II, but since also $\Gamma, A|A \vdash_I \perp$ appending the two deductions together yields a deduction that $\Gamma \vdash_I \perp$.

Theorem 9.1.2 If every classical model for Γ is a model for A then $\Gamma \vdash_C A$.

> **Proof:** We have shown above that $\Gamma \nvdash_I \perp$ iff it has a classical model, such a model is obtained by defining truth valuations on a maximal consistent set containing Γ. So if Γ has no classical model then $\Gamma \vdash_I \perp$ and hence $\Gamma \vdash_C \perp$. To prove completeness for *Clas* with respect to classical models we must show that $\Gamma \vdash_C A$ if and only if $\Gamma, A|A \vdash_C \perp$.[3]

[2] M is consistent such that for every A either $A \in M$ or $M \cup \{A\}$ is inconsistent.

[3] For if every (classical) model for Γ is a model for A then there is no model for $\Gamma \cup \{A|A\}$. But then $\Gamma, A|A \vdash_C$ and so $\Gamma \vdash_C A$.

The left to right direction has already been shown. If $\Gamma, A|A \vdash_C \perp$ then $\Gamma \vdash_C (A|A)|(A|A)$ by the introduction rule. Now

$$\dfrac{(A|A)|(A|A) \quad \cancel{A|A}^1}{A}\; CE(1)$$

is a deduction of A from $(A|A)|(A|A)$ using the rule CE. Therefore if $\Gamma \vdash_C (A|A)|(A|A)$ then $\Gamma \vdash_C A$. ∎

9.1.2 A deduction of $\perp$ can be normalised

Lemma 9.1.1 $\Gamma \vdash_C \perp$ then $\Gamma \vdash_I \perp$.

Proof: I show how to convert a deduction that $\Gamma \vdash_C \perp$ into a deduction that $\Gamma \vdash_I \perp$:[4]

1. Relabel all instances of CI as II.
2. Take any instance of CE in a deduction that $\Gamma \vdash_C \perp$ involving empty discharge, and relabel it IE.
3. Take any other instance of CE and rewrite it as an elimination and subsequent introduction

$$\dfrac{\dfrac{A|B \qquad \begin{array}{c}\cancel{C|C}^n\\ \vdots \\ A\end{array} \qquad \begin{array}{c}\cancel{C|C}^n\\ \vdots \\ B\end{array}}{\perp}\; CE}{(C|C)|(C|C)}\; CI(n)$$

so that $\perp|\perp$ is (empty) discharged in the elimination rule and then $C|C$ is discharged in a subsequent introduction rule.
4. The Prawitz tree has changed so that a formula C has become $(C|C)|(C|C)$, so the Prawitz tree may no longer be a deduction.[5]

[4] A semantic proof that this can be done is simple enough. We proved above that if $\Gamma \nvdash_I \perp$ then maximal consistent sets including Γ are classical models when truth is interpreted as membership. Now, if $\Gamma \vdash_C \perp$ then it has no classical models (by soundness of *Clas*) and so Γ is included in no maximal consistent sets which are classical models when truth is interpreted as membership. Therefore if $\Gamma \vdash_C \perp$ then $\Gamma \vdash_I \perp$.

[5] For example:

$$\dfrac{A|B \qquad \begin{array}{c}\cancel{(C \wedge D)|(C \wedge D)}^n\\ \vdots \\ A\end{array} \qquad \begin{array}{c}\cancel{(C \wedge D)|(C \wedge D)}^n\\ \vdots \\ B\end{array}}{\dfrac{C \wedge D}{D}\; \wedge E}\; CE(n)$$

We must convert the remaining Prawitz tree containing this new formula $(C|C)|(C|C)$ into a deduction of $\perp$.

We need to show that the double Sheffer stroke of the conclusion of each rule is deducible from the double Sheffer stroke of its premises. All cases are considered on pages 121 - 124.

5. Repeat these for all instances of CE in the deduction to obtain a new deduction where all instances of CE involve empty discharge and have been relabelled CI. This is, we have obtained a new deduction using only the rules of *Int*.

So, if $\Gamma \vdash_C \perp$ then applying the method above repeatedly we obtain a deduction that $\Gamma \vdash_I \perp$. ∎

Lemma 9.1.2 A deduction in *Int* may be normalised

Proof: Suppose that $\perp$ is introduced and then eliminated:[6]

$$\dfrac{\dfrac{A|B \quad A \quad B}{\perp}\,IE}{C}\,\perp E$$

then we may remove the unnecessary $\perp$ by inferring C directly with the rule IE.

Now suppose $A|B$ is introduced and then eliminated:

$$\dfrac{\dfrac{\begin{matrix}\cancel{A}^n \quad \cancel{B}^n \\ \vdots \\ \perp\end{matrix}}{A|B}\,II(n) \quad A \quad B}{C}\,IE$$

gets converted to

$$\dfrac{\dfrac{\dfrac{A|B \quad \begin{matrix}\cancel{C \wedge D|C \wedge D}^n \\ \vdots \\ A\end{matrix} \quad \begin{matrix}\cancel{C \wedge D|C \wedge D}^n \\ \vdots \\ B\end{matrix}}{\perp}\,IE}{(C \wedge D|C \wedge D)|(C \wedge D|C \wedge D)}\,II(n)}{D}\,\wedge E$$

and the final step is not a legitimate rule application. We must show that $(D|D)|(D|D)$ is deducible from $(C \wedge D|C \wedge D)|(C \wedge D|C \wedge D)$.

[6]The rule IE is the only way of introducing $\perp$, unless $\perp$ is an assumption, in which case there is a one step normalised deduction of everything.

then we may remove the extraneous introduction of $A|B$ and conclude that $\perp$ directly from A and B (which we used to eliminate $A|B$). We may then apply $\perp E$ to obtain a deduction of C:

$$\dfrac{\begin{array}{c} A \quad B \\ \vdots \\ \perp \end{array}}{C}$$

Also if $A \wedge B$ is introduced and eliminated:

$$\dfrac{\dfrac{A \quad B}{A \wedge B}}{A}$$

then this may be replaced by the deduction of the premise of the elimination rule A (similarly with B).

Finally, if $A \wedge B$ or $A|B$ is introduced by IE or $\perp E$ and then eliminated, then we may infer the conclusion of the second elimination rule directly from the first, for example:

$$\dfrac{\dfrac{\perp}{A \wedge B}\ \perp E}{A}\ \wedge E$$

may be replaced by

$$\dfrac{\perp}{A}\ \perp E$$

Normalisation of *Int* then follows easily. Take any deduction and choose any formula of highest degree[7] that is the conclusion of an introduction rule and is not the conclusion of the deduction. This formula must be eliminated next step,[8] in which case as shown above, the introduction and elimination can be removed. This procedure may be repeated until there are no such formulae, in which case the deduction is normalised. ∎

[7]Not a subformula of any other formula in the deduction.

[8]If it is the minor premise of IE then it is not of maximal degree.

Clas does not so obviously normalise, for suppose a formula is introduced and then eliminated:

$$\dfrac{\dfrac{\begin{array}{c}\cancel{A}^n \quad \cancel{B}^n\\ \vdots\\ \bot\end{array}}{A|B}\,II(n)\qquad \begin{array}{c}\cancel{C|C}^m\\ \vdots\\ A\end{array}\qquad \begin{array}{c}\cancel{C|C}^m\\ \vdots\\ B\end{array}}{C}\,CE(m)$$

We must find a direct deduction of C that we can use to replace this. First we may remove the introduction to obtain a deduction of $\bot$ from $C|C$.

$$\begin{array}{c}\begin{array}{cc}C|C & C|C\\ \vdots & \vdots\\ A & B\end{array}\\ \vdots\\ \bot\end{array}$$

9.1.3 A deduction of $\bot$ from not-A requires a deduction of A

Lemma 9.1.3 If there is a normalised deduction that $\Gamma, C|C \vdash_C \bot$ then there is a normalised deduction that $\Gamma \vdash_C C$.

> **Proof:** I argue by induction on the degree of C. The induction hypothesis is that if there is a normalised deduction that $\Gamma, C|C \vdash_C \bot$ then there is a normalised deduction that $\Gamma \vdash_C C$.
>
> Remember that $\Gamma \vdash_C \bot$ iff $\Gamma \vdash_I \bot$ (theorem 9.1.1), so we may assume that no application of CE occurs in the deduction.
>
> 1. If C is $D|E$ then since
>
> $$\dfrac{\dfrac{\cancel{D|E}^n \quad D \quad E}{\bot}\,IE}{(D|E)|(D|E)}\,II(n)$$
>
> is a deduction of $C|C$ from $\{D, E\}$ it follows that $\Gamma, D, E \vdash_I \bot$ (which can be normalised) and hence $\Gamma \vdash_I C$ with an application of the introduction rule, and so there is a deduction that $\Gamma \vdash_C C$ in normal form.
>
> 2. If C is $A \wedge B$ then we have that
>
> $$\dfrac{\dfrac{A|A \qquad \dfrac{\cancel{A \wedge B}^n}{A}\,\wedge E}{\bot}\,IE}{(A \wedge B)|(A \wedge B)}\,II(n) \qquad \dfrac{\dfrac{B|B \qquad \dfrac{\cancel{A \wedge B}^n}{B}\,\wedge E}{\bot}\,IE}{(A \wedge B)|(A \wedge B)}\,II(n)$$

are deductions of $C|C$ from $A|A$ and from $B|B$. We use these to obtain deductions that $\Gamma, A|A \vdash_I \perp$ and $\Gamma, B|B \vdash_I \perp$. By induction hypothesis there are normalised deductions that $\Gamma, \vdash_I A$ and $\Gamma, \vdash_I B$ and so by $\wedge I$ there is a normalised deduction that $\Gamma, \vdash_I A \wedge B$

3. If C is atomic then the deduction may be changed to a deduction of C by adding an application of $\perp E$, furthermore the deduction of C from $\Gamma \cup \{C|C\}$ may be normalised as it contains no application of CE.

   There are two cases to consider:

   (a) Suppose element of Γ contains an occurrence of Sheffer stroke. I show that for any deduction in normal form of atomic C from assumptions $\Gamma \cup \{C|C\}$ (Γ not containing Sheffer stroke) there is a deduction in normal form of C from Γ. I argue by induction on the length of the deduction of C from $\Gamma \cup \{C|C\}$. If the deduction involves no inference rules then C is one of the assumptions and there is nothing to prove. Otherwise, since C is atomic the final step of the deduction is an elimination rule the major premise of which is either an assumption or is itself the conclusion of an elimination rule. This is because the deduction is in normal form. Thus C is the conclusion of the last of a sequence of elimination rules, the conclusion of each is the major premise of the next;[9] the major premise of the first, say P, is an assumption.[10]

      i. If P is $C|C$ then it is used in an application of IE (no undischarged assumption is used in an application of CE):

      $$\frac{C|C \quad C}{X}$$

      the minor premises of which involve a shorter deduction of C from $\Delta \cup \{C|C\}$ where $\Delta \subseteq \Gamma$ and so there is a deduction that $\Delta \vdash_I C$ and thence that $\Gamma \vdash_I C$.[11]

      ii. If P is $A \wedge B$ then it is used to infer A or B. If we delete P then we obtain a shorter deduction that A or B

[9] This is because the deduction is in normal form.

[10] For there is no introduction rule below to discharge it.

[11] The minor premises can contain no more assumptions than $\Gamma \cup \{C|C\}$ as there are no introduction rules below them to make discharges.

is an assumption where P used to be. That is, we have a deduction that $\Delta \cup \{C|C\} \vdash_I C$ where $\Gamma = \Delta \cup \{A\}$ or $\Gamma = \Delta \cup \{B\}$. By induction hypothesis there is a deduction in normal form that $\Delta \vdash_I C$ and since $\Gamma \vdash_I A$ and $\Gamma \vdash_I B$ with the application of an elimination rule, there is a deduction that $\Gamma \vdash_I C$.

(b) Suppose that C is atomic and some formula in Γ contains an occurrence of Sheffer stroke.
Then successive applications of $\wedge E$ will deduce some $D|E$ from Γ, then $\Gamma, C|C \vdash_I \perp$. In particular $\Gamma, C|C \vdash_I E$ and $\Gamma, C|C \vdash_I F$, and so there are normalised deductions of D and E from $\{\Gamma, C|C\}$ which we may use to eliminate $D|E$ to conclude that C.

$$\dfrac{D|E \qquad \begin{matrix}\cancel{C|C}^n\\ \vdots \\ D\end{matrix} \qquad \begin{matrix}\cancel{C|C}^n\\ \vdots \\ E\end{matrix}}{C}\ CE(n)$$

The deduction is still normalised as $\Gamma \vdash_C D|E$ by elimination rules alone and so was not introduced by an introduction rule.

∎

Theorem 9.1.3 Deductions in *Clas* may be normalised.

Proof: Any deduction that $\Gamma \vdash_C A$ may be used to produce a deduction that $\Gamma \cup \{A|A\} \vdash_C \perp$ which can be normalised, and therefore, by the lemma above there is a normalised deduction that $\Gamma \vdash_C A$. ∎

9.1.4 Remarks

Looking at the proof we see that *Clas* does not have the subformula property. However it comes close, if $\Gamma \vdash_C A$ then there is a deduction that uses at most subformulae of (a formula in) Γ and A and also $\perp$ and $C|C$ where C is an *atomic* subformula of Γ and A.

9.2 The Quantified case

9.2.1 The universal quantifier

Against the spirit of Sheffer stroke, but for simplicity (and plausibility) I shall add rules for a universal quantifier as well.

$$\frac{\forall x A}{A[x/t]}\ \forall E \qquad \frac{A}{\forall x A[y/x]}\ \forall I$$ provided A does not depend on any formulae in which y is free.

Let $Clas_1$ be the logic obtained by adding the quantifier rules above and the rule IE to $Clas$. Let $\Gamma \vdash_{C_1} A$ mean that there is a deduction in $Clas_1$ the assumptions of which are in Γ and with the conclusion A.

9.2.2 Tree conversions

Lemma 9.2.1 From any deduction that $\Gamma \vdash_{C_1} \perp$, we may obtain a deduction that $\Gamma \vdash_{C_1} \perp$ in which the major premise of every application of CE is at a top node.

Proof:

Step 1 Take any instance of CE for which the major premise $A|B$ is not at a top node in the deduction tree, and rewrite it as an elimination and subsequent introduction with CI

$$\cfrac{\cfrac{A|B \qquad \begin{matrix}\cancel{C|C}^n \\ \vdots \\ A\end{matrix} \qquad \begin{matrix}\cancel{C|C}^n \\ \vdots \\ B\end{matrix}}{\perp}\ IE}{(C|C)|(C|C)}\ CI(n)$$

so that $\perp|\perp$ is (empty) discharged in the elimination rule and then $C|C$ is discharged in a subsequent introduction rule.

Step 2 The Prawitz tree now has been changed so that a formula C has become $(C|C)|(C|C)$, and therefore it may not be a deduction.[12]

(a) If C is $\perp$ then:

$$\cfrac{(\perp|\perp)|(\perp|\perp) \qquad \cfrac{\cancel{\perp}^n}{\perp|\perp}\ II}{\perp}\ IE(n)$$

[12] For an example see footnote 5 on page 115.

appended to the deduction of $(C|C)|(C|C)$ yields a deduction of $\bot$ from $(C|C)|(C|C)$, and the deduction may then continue as before (as we deduced C from $(C|C)|(C|C)$).

(b) If C was a premise in an application of $\forall E$, $\wedge E$, $\wedge I$, or IE then the original inference step had the form

$$\frac{F \quad G}{D}\ \forall E, \wedge I, \wedge E, IE$$

where one of $\{F, G\}$ is C. Let us suppose that C is G and that the new, illegitimate step is:

$$\frac{F \quad (C|C)|(C|C)}{D}$$

then we may replace this by the following deduction of $(D|D)|(D|D)$ from F and $(C|C)|(C|C)$:

$$\dfrac{\dfrac{(C|C)|(C|C) \qquad \dfrac{\dfrac{\cancel{D|D}^m \qquad \dfrac{F \quad \cancel{C}^n}{D}\ \forall E, \wedge I, \wedge E, IE}{\bot}\ IE}{C|C}\ II(n)}{\bot}\ IE}{(D|D)|(D|D)}\ II(m)$$

(c) If C was a premise in an application of CE which has now become:

$$\dfrac{(C|C)|(C|C) \qquad \begin{matrix}\cancel{D|D}^m\\ \vdots\\ X\end{matrix} \qquad \begin{matrix}\cancel{D|D}^m\\ \vdots\\ Y\end{matrix}}{D}\ CE(m)$$

where C is $E|F$. Then we may convert it into a deduction of $(D|D)|(D|D)$ as in case (b) where E and F are taken as deduction premises:

$$\dfrac{\dfrac{(C|C)|(C|C) \qquad \dfrac{\dfrac{\cancel{D|D}^m \qquad \dfrac{\begin{matrix}\cancel{D|D}^m\\ \vdots\\ X\end{matrix} \qquad \begin{matrix}\cancel{D|D}^m\\ \vdots\\ Y\end{matrix} \qquad \cancel{C}^n}{D}\ CE}{\bot}\ IE}{C|C}\ II(n)}{\bot}\ IE}{(D|D)|(D|D)}\ II(m)$$

On the other hand, if C was the conclusion of one of the minor premises:

$$\dfrac{A|B \qquad \begin{array}{c}\cancel{D|D}^m\\ \vdots\\ C\end{array} \qquad \begin{array}{c}\cancel{D|D}^m\\ \vdots\\ B\end{array}}{D}\; CE(m)$$

(where C is A) then the problematic inference step is of the form:

$$\dfrac{A|B \quad \begin{array}{c}\cancel{D|D}^m\\ \vdots\\ (C|C)|(C|C)\end{array} \qquad \begin{array}{c}\cancel{D|D}^m\\ \vdots\\ B\end{array}}{D}\; CE(m)$$

which may be replaced by:

$$\cfrac{\cfrac{A|B \qquad \cfrac{\begin{array}{c}\cancel{D|D}^m\\ \vdots\\ (C|C)|(C|C)\end{array} \qquad \cfrac{\cfrac{A|B \quad \cancel{C}^n \quad \begin{array}{c}\cancel{D|D}^m\\ \vdots\\ B\end{array}}{\bot}\; IE}{C|C}\; II(n)}{C}\; IE \qquad \begin{array}{c}\cancel{D|D}^m\\ \vdots\\ B\end{array}}{\bot}\; IE}{(D|D)|(D|D)}\; II(m)$$

(d) If C is the premise to an application of $\forall I$ then the old inference step

$$\dfrac{C}{\forall x C}$$

has been replaced by:

$$\dfrac{(C|C)|(C|C)}{\forall x C}$$

The deduction of $(C|C)|(C|C)$ does not depend on any assumptions in which x is free (for the original inference was legitimate), so we may replace the problematic inference step with:[13]

$$\dfrac{\dfrac{\cancel{\forall xC|\forall xC}^n \qquad \dfrac{\dfrac{\dfrac{\dfrac{\cancel{\forall xC|\forall xC}^n \qquad \dfrac{\begin{array}{c}\vdots\\ (C|C)|(C|C)\end{array} \quad \cancel{C|C}^m}{\forall xC}\,IE}{\perp}\,IE}{C}\,CE(m)}{\forall xC}\,\forall I}{}}{\perp}\,IE}{(\forall xC|\forall xC)|(\forall xC|\forall xC)}\,II(n)$$

where $\forall xC|\forall xC$ assumed, used to deduce C and then used to introduce $(\forall xC|\forall xC)|(\forall xC|\forall xC)$. Although this deduction has added an extra application of CE it is applied to a formula that is at a top node in the Prawitz tree. The extra application of CE eliminates $\forall xC|\forall xC$ which is an assumption, this assumption is soon after discharged by the application of II.

Step 3 After applying step 1 we will have produced a Prawitz tree that is not a deduction of $\perp$ because an occurrence of C has been changed to $(C|C)|(C|C)$ (the Prawitz tree may not be a deduction at all). The new Prawitz tree does, however, have one fewer application of CE (it has been replaced by CI.

(a) If C was the conclusion of the deduction, then C is $\perp$ and step 1 has converted the deduction into a deduction of $(\perp|\perp)|(\perp|\perp)$. After applying step 2 we obtain a deduction of $\perp$.

(b) If C was the premise of a rule application the conclusion of which was D, then after applying step 2 we will have inserted a deduction of $(D|D)|(D|D)$ from $(C|C)|(C|C)$. The new Prawitz tree is now not a deduction of $\perp$ because of $(D|D)|(D|D)$ rather than because of $(C|C)|(C|C)$.

[13] We cannot follow the previous strategy and use the legitimate application of $\forall I$ to obtain:

$$\dfrac{\dfrac{\begin{array}{c}\vdots\\ (C|C)|(C|C)\end{array} \qquad \dfrac{\dfrac{\cancel{\forall xC|\forall xC}^n \quad \dfrac{\cancel{C}^m}{\forall xC}\,\forall I!!}{\perp}\,IE}{C|C}\,II(m)}{\perp}\,IE}{(\forall xC|\forall xC)|(\forall xC|\forall xC)}\,II(n)$$

as the side conditions on the rule application of $\forall I$ are not met when C is separated from the deduction and becomes an assumption.

Since D is one inference step closer to the conclusion we should keep reapplying step 2 until we obtain a deduction of $\perp$.

Step 4 After applying steps 1–3 we obtain a deduction of $\perp$ with one fewer application of CE to a formula that is not at a top node. Therefore, we should make successive applications of steps 1–3 until we remove all such occurrences.

∎

Lemma 9.2.2 Any deduction that $\Gamma \vdash_{C_1} \perp$ may be normalised.

Proof: If $\forall x A$ is introduced and eliminated:

$$\dfrac{\dfrac{A}{\forall x A[y/x]}\ \forall I}{A[y/t]}\ \forall E$$

then because of the condition on $\forall I$ we may replace y by t in the deduction of A to obtain a more direct deduction of $A[y/t]$ (which will replace the introduction and subsequent elimination). The reduction cases for the other connectives are as in lemma 9.1.2.

Now, take any deduction that $\Gamma \vdash_{C_1} \perp$, by lemma 9.2.1 it may be converted into a deduction where the major premises of all applications of CE are formulae at top nodes. Now choose any formula of highest degree[14] that is the conclusion of an introduction rule and is not the conclusion of the deduction. This formula must be eliminated next step, and is not eliminated by CE (it may be eliminated by IE instead), in which case the introduction and elimination can be removed. This procedure may be repeated until there are no such formulae, in which case the deduction is normalised. ∎

9.2.3 A deduction of $\perp$ from not-A requires a deduction of A

Lemma 9.2.3 From a deduction of $\perp$ from $\{C|C\} \cup \Gamma$ there is a normalised deduction of C from assumptions Γ.

Proof: I argue by induction on the degree of C that if there is a normalised deduction that $\Gamma, C|C \vdash_{C_1} \perp$ then there is a normalised deduction that $\Gamma \vdash_{C_1} C$.

Given the lemmas of the previous section we may convert the deduction of $\perp$ from $\{C|C\} \cup \Gamma$ into a normalised deduction that $\{C|C\} \cup \Gamma \vdash_{C_1} \perp$.

[14] Not a subformula of any other formula in the deduction.

1. The cases where C is $D|E$ or $A \wedge B$ are considered in lemma 9.1.3 on page 118.

2. If C is $\forall x A$ then choosing y to be a variable that does not occur free in Γ

$$\dfrac{\dfrac{(A|A)[x/y] \quad \dfrac{\cancel{\forall x A}}{A[x/y]}\ \forall E}{\bot}\ IE}{\forall x A | \forall x A}\ CI$$

   is a deduction of $C|C$ from $(A|A)[x/y]$, so there is a deduction that $\Gamma, (A|A)[x/y] \vdash_{C_1} \bot$ which can be normalised and by induction hypothesis there is a normalised deduction that $\Gamma \vdash_{C_1} A[x/y]$. Since y is not free in Γ and Γ contains all undischarged assumptions we may apply $\forall I$ to obtain a normalised deduction that $\Gamma \vdash_{C_1} \forall x A$.

3. The case where C is atomic is for the most part considered in lemma 9.1.3. What is missing is one extra subcase where P is $\forall x A[t/x]$:

   iii. If P is $\forall x A[t/x]$ then it is used to deduce A. We may then delete P to obtain a shorter deduction of C from $\Delta \cup \{C|C\}$ where $\Delta = \Gamma \cup \{A\}$.[15] By induction hypothesis there is a deduction of C from Δ in normal form and since A follows from Γ by an elimination rule there is a normalised deduction of C from Γ.

∎

Theorem 9.2.1 Every deduction in *Clas*$_1$ can be reduced to a deduction in normal form that does not use IE.

Proof: Take any deduction that $\Gamma \vdash_{C_1} A$, we may use this to produce a deduction that $\Gamma \cup \{A|A\} \vdash_{C_1} \bot$ from which, by lemma 9.2.3, we can obtain a normalised deduction that $\Gamma \vdash_{C_1} A$. Furthermore, we may re-label all applications of IE in this deduction as CE to obtain a new deduction that $\Gamma \vdash_{C_1} A$ (see the proof of theorem 9.1.1). ∎

[15] Since no introduction rules depend on P we need not worry about t occurring free in the assumption replacing P.

9.2.4 Soundness and Completeness

We should give the universal quantifier its usual interpretation: $v(\forall xA)$=T iff $v'(A) = T$ for every valuation v' x-alternate to v. On doing this a soundness and completeness proof of $Clas_1$ for the familiar first order semantics is amusing enough, but not necessary here. It is enough to notice that the addition of familiar quantifier rules to a deduction system that is complete for classical propositional semantics should yield first order system that is complete for classical first order semantics.

9.3 Concluding remarks

We can now argue that the true basis for logic is incompatibility (Sheffer stroke), universal quantification and conjunction.[16] This basis has autonomously defined and harmonious introduction and elimination rules. We may use this basis to define other connectives. For example, I think it quite plausible that we define negation in terms of sheffer stroke using these rules:

$$\dfrac{\begin{matrix}\cancel{A}^n\\ \vdots\\ \bot\end{matrix}}{\sim A}\ \sim I(n) \qquad\qquad \dfrac{\sim A \qquad \begin{matrix}\cancel{B|B}^n\\ \vdots\\ A\end{matrix}}{B}\ \sim E(n)$$

and then disjunction in terms of negation using these rules:

$$\dfrac{A}{A \vee B}\ \vee IIa \qquad \dfrac{B}{A \vee B}\ \vee IIb \qquad \dfrac{\begin{matrix}\cancel{\sim A}^n \quad \cancel{\sim B}^n\\ \vdots\\ \bot\end{matrix}}{A \vee B}\ \vee I(n) \qquad \dfrac{A \vee B \qquad \begin{matrix}\cancel{A}^n\\ \vdots\\ C\end{matrix} \qquad \begin{matrix}\cancel{B}^n\\ \vdots\\ C\end{matrix}}{C}\ \vee E(n)$$

Thus it may seem that disjunction and negation are basic logical connectives, but in fact they are defined in terms of the a true basic connective sheffer stroke.

As the lengths of the chapters suggest, the restart rule provides a simpler solution to the problem of normalisation for classical logic. In general I find it more elegant and it is the rule I shall adopt. I think it likely that we have restart present in the structure of our deduction and then define our connectives by the usual rules. I think it is less likely that sheffer stroke is our basic connective (i.e. that incompatibility is our basic logical concept) and other connectives are defined in terms or with reference to it. However I do not find it completely implausible that this turn out to be the case.

[16] See [?] for an argument that negation derives from a primitive understanding of incompatibility, it just so happens that incompatibility does far more than yield negation.

9.3.1 The conclusion

So, in conclusion, I think there is a strong case to be made that sheffer stroke, conjunction and the universal quantifier are the basic logical connectives. But the case is not strong enough, the case for the restart rule is stronger as restart is more general and more simple. Consequently I maintain the thesis that are basic logic consists of the familiar introduction and elimination rules for the logical connectives (conjunction, disjunction etc.) and makes use of the restart rule (a form of structural indirect proof).

Chapter 10

Strict implication

In this chapter I attempt to tackle two problems at once. I shall present some rules for a strict conditional that I claim captures the logic of validity, or at least strict implication. I take it that philosophical analysis of validity (in terms of possible worlds) has indicated that the logic of strict implication is at least $S4$. In addition to this, I claim to have resolved a familiar problem relating to the addition of intuitionistic implication to classical logic, I outline the problem in section 10.1.1.

10.1 Strict implication

10.1.1 The collapse

It is well known that having classical rules and intuitionistic rules together results in the intuitionistic implication collapsing into material implication. The problem is (letting $\supset$ and $\sim$ be classical and $\rightarrow$ and $\neg$ be intuitionistic) merging the two logics yield that $A \supset B \vdash A \rightarrow B$ and $\sim A \rightarrow \neg A$.[1] Because we can make deductive headway by introducing classical connectives and then turning them into intuitionistic ones by a subsequent elimination, we also lose normalisation.[2] Here is an example, suppose the rules for $\rightarrow$ are added

[1] The deductions

$$\dfrac{\dfrac{\cancel{A}^1 \quad A \rightarrow B}{B}\,{\rightarrow}E}{A \supset B}\,{\supset}I(1) \qquad \dfrac{\dfrac{\cancel{A}^1 \quad A \supset B}{B}\,{\supset}E}{A \rightarrow B}\,{\rightarrow}I(1)$$

show that classical and intuitionistic conditionals become inter-deducible when the rules for them are naively added together without modification.

[2] A natural way of introducing intuitionistic implication to Read's system is to add the new implication $\rightarrow$ and demand a single conclusion for the introduction of $A \rightarrow B$, and

to the sheffer stroke logic *Clas* (the conclusion is a famous non-theorem of intuitionistic logic).

$$\dfrac{\dfrac{\cancel{A|A}^1 \quad \dfrac{\dfrac{\cancel{A|A}^2 \quad \dfrac{\cancel{(A \to B) \to A} \quad \dfrac{\dfrac{\cancel{A|A}^2 \quad \cancel{A}^3}{B}}{A \to B}\,!!(3)}{A}}{\perp}}{(A|A)|(A|A)}\,CI(2)}{A}\,CE(1)}{[(A \to B) \to A] \to A}\,{\to}I$$

There is an introduction and subsequent elimination of sheffer stroke in the deduction. The deduction will not normalise by the method above as above the application of CI we have a deduction of $\perp$ from $A|A$ and $\Gamma = \{(A \to B) \to A\}$ which contains no formula of the form $E|F$, but also contains some non-atomic formulae. Notice also the inference marked '!!', this is an inference of $A \to B$ from $A|A$.[3]

Looking at the introduction rules, we can see why: *they are the same.* In the Sheffer stroke logics *Int* and *Clas* the introduction rules are identical, and so are the introduction rules for $\to$ and $\supset$ with or without restart.

The problem exemplified by ! is this: $A \to B$ suggests a necessary connection between A and B. $A \to B$ should not be true merely by accident but because of a strict relationship between A and B. $A \supset B$ on the other hand does not imply such a necessary connection, $A \supset B$ may be true only because of an accidental relationship between A and B. For example if B happens to be true then $A \supset B$ is true, but $A \to B$ should not be true unless there is a stronger link between A and B.

10.1.2 The nature of the collapse

Here are some intuitive inference rules for strict implication, notice the similarity for the rules for the conditional in intuitionistic logic.

$$\dfrac{A \quad A \to B}{B} \qquad \dfrac{\begin{matrix}\cancel{A} \\ \vdots \\ B\end{matrix}}{A \to B}$$

for Rumfitt's to give only the rules for $+A \to B$. Neither of these work.

[3] The system with restart suffers a similar collapse, the collapse is more obvious as the new rules for $\to$ are identical to the rules for $\supset$. This entails that normalisation does not fail (though intuitionistic and classical implications become indistinguishable).

A major problem is then that these seemingly indisputable inference rules (when taken as the essence of the natural language conditional) yield strange results.

- We may infer $B \rightarrow A$ from A

- We may infer $B \rightarrow A$ from $\sim B$

The two results together constitute the least enjoyable part of teaching an elementary logic course to philosophy students, they do not seem right for a strict implication.

The problem is that the introduction rule above is misleading. Perhaps a less misleading way of putting it is this:

$$\frac{\begin{array}{c}\cancel{A}, \Gamma \\ \vdots \\ B\end{array}}{A \rightarrow B}$$

where Γ contains all the assumptions used in the inference of B from A. The rule for the conditional allows the inference B from A on the basis of anything that can be put in Γ. The classical inference rule places little restriction on what can be a member in Γ. Anything we know to be true can go in Γ, for example if we know that B is true then we can put B in Γ and then the deduction of B from $\{A\} \cup \Gamma$ is easy. When we use a strict conditional, on the other hand, there are restrictions on what can be put in Γ, we try to infer B from A only using truths that do not vary with the context. This is so that $A \rightarrow B$ retains a generality across various contexts. This matches te difference between validity and local validity, one is independent of the context and the other is not. The material conditional expresses the relation of local validity, and we must now give rules for a strict conditional that expresses the relation of validity.

10.2 The addition of strict implication

We will divide formulae into two categories, *strong* and *weak*. Strong formulae are ones which are true or false independently of the context. Weak formulae are not so universal. Effectively, a strong formula such that if it is true then it is necessarily true.

10.2.1 The rules for the strict conditional $\to$

Suppose then that we infer B from Γ and the additional assumption that A. The inference may rely on weak formulae in Γ which are context dependent. So we may not wish to conclude that $A \to B$ for when the context changes Γ is no longer true.[4] This legitimates introducing $A \to B$ on the basis of a deduction of B from $\{\Gamma, A\}$ when the deduction of B from A depends only on A and strong context independent) formulae in or deducible from Γ.

- Any formula of the form $A \to B$ is strong
- $\perp$ is strong and so is $\top$

The following rules for $\to$ a to extend a natural deduction system for classical logic which already contains the connectives $\wedge, \perp, \supset$ etc.

The elimination rule for $\to$ remains Modus Ponens

$$\frac{A \quad A \to B}{B} \to E$$

but the introduction rule gets an extra side condition

$$\frac{C_1 \ldots C_n \qquad \begin{array}{c} \cancel{A}^m, \cancel{C_1}^m \ldots \cancel{C_n}^m \\ \vdots \\ B \end{array}}{A \to B} \to I(m)$$

provided that the C_i are all strong, and the inference of B depends on no formulae other than $A, C_1 \ldots C_n$ and depends on no weak rule applications.[5]

The inference of B from $A, C_1 \ldots C_n$ must not contain on any weak rule applications, and so it must be a deduction: any application of restart on which B depends is weak, so all applications of restart on which B depends must be completed at B.[6] A special case of $\to I$ is when n is 0 (i.e. there

---

[4] At most we can conclude that if A is true *as well as* the weak assumptions on which it relied, then so is B.

[5] So far the restart rule is the only rule I have used that can be weak. See section 11.3 for some more rules that are always weak.

[6] Here is the corresponding sequent rule $\to$*right*, note its similarity to the CUT rule but remember that $\to$*right*, unlike CUT, always increases complexity.

$$\frac{\Gamma \vdash C_1 \quad \ldots \quad \Gamma \vdash C_n \quad \{C_1 \ldots C_n, A\} \vdash B}{\Gamma \vdash A \to B} \to right$$

provided that the C_i are all strong

The side condition is not quite complete as I have provided no way of identifying assumptive rules in a sequent calculus. In this thesis I shall not show how to incorporate assumptive rules into a sequent calculus.

are no C_i), in such a case the inference rule has no premises other that the deductive premise which is a deduction of B from A alone. For example in this deduction of $A \to A$ from the emptyset:

$$\frac{\cancel{A}^1}{A \to A} \to I(1)$$

In the case where there are no C_i, $\to I$ becomes the more familiar conditional proof with the restriction that the deduction of B from A depends on no assumption other than A. If $A \supset B$ is a theorem then B may be deduced from A depending on no other assumptions. We can use this deduction applying $\to I$ with no C_i to conclude that $A \to B$. Thus if $A \supset B$ is a theorem then so is $A \to B$.

The restriction allows us to introduce $A \to B$ only if B is derived only from A and *strong conclusions* of other possibly weak assumptions.

Theorem 10.2.1 $(A \wedge A_1 \wedge \cdots \wedge A_n) \to B$ is logically equivalent to $A \to (A_1 \supset \cdots \supset A_n \supset B)$.

Proof: This:

$$\dfrac{\dfrac{(A \wedge A_1 \wedge \cdots \wedge A_n) \to B \qquad \begin{array}{c} A, \cancel{A_1}^1, \ldots, \cancel{A_n}^n \\ \vdots \ \wedge I \\ A \wedge A_1 \wedge \cdots \wedge A_n \end{array}}{B}\to E}{\begin{array}{c} A_n \supset B \\ \vdots \ \supset I(2) \ldots (n) \\ A_1 \supset \ldots A_n \supset B \end{array}} \supset I(1)$$

is a deduction of $A_1 \supset \ldots A_n \supset B$ depending only on A and a strict implication $(A \wedge A_1 \wedge \cdots \wedge A_n) \to B$ (which is strong) and so we may finish the deduction with an $\to$ introduction (let $m > n$):[7]

$$\dfrac{(A \wedge A_1 \wedge \cdots \wedge A_n) \to B \qquad \dfrac{\dfrac{(A \wedge A_1 \wedge \cdots \wedge A_n) \to B \qquad \begin{array}{c} \cancel{A}^m, \cancel{A_1}^1, \ldots, \cancel{A_n}^n \\ \vdots \ \wedge I \\ A \wedge A_1 \wedge \cdots \wedge A_n \end{array}}{B}\to E}{\begin{array}{c} A_n \supset B \\ \vdots \ \supset I(2) \ldots (n) \\ A_1 \supset \ldots A_n \supset B \end{array}} \supset I(1)}{A \to (A_1 \supset \cdots \supset A_n \supset B)} \to I(m)$$

[7]Treating $(A \wedge A_1 \wedge \cdots \wedge A_n) \to B$ as the only C_i.

For the other half of the equivalence, from $A \to (A_1 \supset \cdots \supset A_n \supset B)$ we may apply modus ponens successively

$$\dfrac{\dfrac{A \wedge A_1 \wedge \cdots \wedge A_n}{A_n}\,\wedge E \qquad \begin{array}{c}\dfrac{\dfrac{A \wedge A_1 \wedge \cdots \wedge A_n}{A}\,\wedge E \quad A \to (A_1 \supset \cdots \supset A_n \supset B)}{A_1 \supset \cdots \supset A_n \supset B}\,\to E \\ \vdots\ \supset E, \wedge E \\ A_n \supset B\end{array}}{B}\,\supset E$$

to obtain a deduction of B from the premises $A \to (A_1 \supset \cdots \supset A_n \supset B)$ (which is strong) and $A \wedge A_1 \wedge \cdots \wedge A_n$. So we may discharge $A \wedge A_1 \wedge \cdots \wedge A_n$ for a final introduction of $\to$.

$$\dfrac{A \to (A_1 \supset \cdots \supset A_n \supset B) \qquad \dfrac{\dfrac{\cancel{A \wedge A_1 \wedge \cdots \wedge A_n}^{1}}{A_n}\,\wedge E \qquad \begin{array}{c}\dfrac{\dfrac{\cancel{A \wedge A_1 \wedge \cdots \wedge A_n}^{1}}{A}\,\wedge E \quad \cancel{A \to (A_1 \supset \cdots \supset A_n \supset B)}^{1}}{A_1 \supset \cdots \supset A_n \supset B}\,\to E \\ \vdots\ \supset E, \wedge E \\ A_n \supset B\end{array}}{B}\,\supset E}{(A \wedge A_1 \wedge \cdots \wedge A_n) \to B}\,\to I(1)$$

∎

As a taster of how I shall explain away some problems with embedded conditionals. We may analyse

(†) if A then if B then C

as

$$A \to (B \supset C)$$

which is equivalent to

$$(A \wedge B) \to C$$

which is the analysis of

(‡) if A and B, then C.

It is natural to consider it a failure of a theory of conditionals if † and ‡ are not equivalent. However, I shall not argue that the natural language conditional is the strict conditional or the material conditional (or some combination of the two).

10.2.2 Semantics of the strict conditional

Let $\mathcal{L}$ be classical logic, the argument below is not dependent on which system of rules is used to formulate $\mathcal{L}$ however in this thesis the received formulation of classical logic is the system on page 87 using the restart rule. Let $\mathcal{L}_i$ be classical logic with the new implication rules for $\rightarrow$ and the following inference rule for $\top$ is added,

$$\frac{}{\top}\ \top I$$

and also with these additional rules for strength above.

Let M_L is a set of models of $\mathcal{L}$ and $\langle W, R\rangle$ is an $S4$ (transitive and reflexive) Kripke frame, $\langle W, R, M_L, f\rangle$ is a model of $\mathcal{L}_i$ where f is a bijection from W to M_L.[8] Truth is then defined as follows, let M be a model of $\mathcal{L}_i$ and suppose $w \in W$

- For any w in any model M, $M \vDash_w \top$ and $M \nvDash_w \bot$.
- If A is atomic then $M \vDash_w A$ iff $f(w) \vDash A$ (remember that $f(w) \in M_L$)
- If A is $B \rightarrow C$ then $M \vDash_w A$ iff $M \vDash_{w'} C$ for all w' s.t. wRw' and $M \vDash_{w'} B$
- If A is $B \wedge C$ then $M \vDash_w A$ iff both $M \vDash_w B$ and $M \vDash_w C$
- If A is $B \vee C$ then $M \vDash_w A$ iff either $M \vDash_w B$ or $M \vDash_w C$
- If A is $B \supset C$ then $M \vDash_w A$ iff either $M \nvDash_w B$ and $M \vDash_w C$

If a formula A is strong (i.e. a strict conditional) then at any world w in any model, either A is not true at w or it is true at all w' s.t. wRw'

Now we may prove soundness for this semantics:

Theorem 10.2.2 If $\Gamma \vdash A$ in $\mathcal{L}_i$ then A is true every model of $\mathcal{L}_i$ where Γ is true.

> **Proof:** *The proof I present here is for a logic where* restart *is replaced by PIP. See section 8.2 to see how* restart *and PIP are interchangeable (i.e. switching one for the other does not alter the logical consequence relation). In the following proof I omit the clause for PIP.*[9]

[8] A model for $S4$ consists of a set of possible worlds with a transitive and reflexive accessibility relation on them. $A \rightarrow B$ is true at a world in wuch a model if B is true at every accessible world where A is true.

[9] Remember that when restart is replaced by PIP we need not worry about assumptive rule applications (as incomplete applications of restart are the only assumptive rule applications that appear in this thesis).

We must show, by induction on the length of the derivation, that for any world w if $\Gamma \vdash_{\mathcal{L}_i} C$ and Γ is true at w then C is true at w. It is a simple if somewhat tedious matter to check all the rules. The only interesting case is $\rightarrow I$, but I consider all cases here.

Firstly if $A \rightarrow B$ is true at w then it is true at all accessible w', thus the condition on strength is sound.

If $C \in \Gamma$ then the result is immediate.

If C is derived by $\wedge E$ then we have a shorter deduction that $\Gamma \vdash_{\mathcal{L}_i} C \wedge D$ (or $D \wedge C$) for some D which by induction hypothesis is true at w and so C is true at w. If C is $A \wedge B$ and is derived by $\wedge I$ then by induction hypothesis A and B are true at w virtue of the shorter deductions that $\Gamma \vdash_{\mathcal{L}_i} A$ and $\Gamma \vdash_{\mathcal{L}_i} B$, and therefore so is C.

If C is derived by $\vee E$ then we have shorter deductions that $\Gamma \vdash_{\mathcal{L}_i} A \vee B$, $\Gamma, A \vdash_{\mathcal{L}_i} C$, $\Gamma, B \vdash_{\mathcal{L}_i} C$. By induction hypothesis $A \vee B$ is true at w and so one of A or B is true at w and therefore, also by induction hypothesis, so is C. If C is $A \vee B$ and is derived by $\vee I$ then by induction hypothesis A or B is true at w virtue of a shorter deduction either that $\Gamma \vdash_{\mathcal{L}_i} A$ or that $\Gamma \vdash_{\mathcal{L}_i} B$, and therefore so is C.

If C is derived by $\supset E$ then we have shorter deductions that $\Gamma \vdash_{\mathcal{L}_i} D \supset C$ and $\Gamma \vdash_{\mathcal{L}_i} D$ for some D. By induction hypothesis D and $D \supset C$ are both true at w and then so is C. If C is $A \supset B$ and is derived by $\supset I$ then virtue of the shorter deduction that $\Gamma, A \vdash_{\mathcal{L}_i} B$ and the induction hypothesis either A is not true at w or it is true together with Γ and so is B.

$\bot$ is not true at any world, so $\bot E$ is a sound rule for this semantics, and $\top$ is true at all worlds so $\top I$ is sound.

If C is derived by $\rightarrow E$ then we have shorter deductions that $\Gamma \vdash_{\mathcal{L}_i} D \rightarrow C$ and $\Gamma \vdash_{\mathcal{L}_i} D$ for some D. By induction hypothesis D and $D \rightarrow C$ are both true at w and then, since wRw, so is C. If C is $A \rightarrow B$ and is derived by $\rightarrow I$ then there is some set of strong formulae Δ such that $\Gamma \vdash_{\mathcal{L}_i} D$ for all $D \in \Delta$ and $\Delta, A \vdash_{\mathcal{L}_i} B$ (all of which are shorter deductions). By induction hypothesis Δ is true at w, and since every member of Δ is strong Δ is true at all w' s.t. wRw'. Thus for any w' s.t. wRw' at which A is true, both A and Δ are true, and by induction hypothesis so is B. ∎

10.2.3 Completeness

Before proving completeness notice that $A \to (B \to A)$ is not a theorem of this logic. We cannot infer $B \to A$ from A as A might not be strong. For example if, in a model, A is true at w and not in all w' s.t. wRw' and B is true in all such w' then A is true at w while $B \to A$ is false.

Also see that if

$$(\dagger) B_1 \dots B_n \vdash_{\mathcal{L}_i} C$$

then

$$(A \to B_1) \dots (A \to B_n) \vdash_{\mathcal{L}_i} (A \to C)$$

the deduction for this is:

$$\dfrac{A \to B_1 \dots A \to B_n \qquad \begin{array}{c}\cancel{(A \to B_1)}^1, \dots, \cancel{(A \to B_n)}^1, \cancel{A}^1 \\ \vdots\ {\to}E \\ B_1 \dots B_n \\ \vdots\ \dagger \\ C\end{array}}{A \to C}\ {\to}I(1)$$

by appending deduction † to the bottom of n applications of modus ponens. The final inference (introduction of $A \to B$) is clearly legitimate because each the $A \to B_i$ is strong and the deduction of C from the B_i depends on no undischarged assumptions.

Theorem 10.2.3 If A is true every model of $\mathcal{L}_i$ where Γ is true, then $\Gamma \vdash A$ in $\mathcal{L}_i$.

> **Proof:** To show this we shall assume that $\Gamma \nvdash A$ and show that there is a model of $\mathcal{L}_i$ where Γ is true but A is not. This is done by constructing a canonical model.
>
> Take all maximal consistent sets of $\mathcal{L}_i$[10] – we denote such a set by m – and define an accessibility relation R on them (we regard each m as a world)
>
> $$mRm' \text{ iff } \{B : A \to B \in m \text{ and } A \in m'\} \subseteq m'$$
>
> We must show that we have a Kripke model. We must prove that
>
> $$A \to B \in m \text{ iff } B \in m' \text{ for all } m' \text{ s.t. } mRm' \text{ and } A \in m'$$

[10] A maximal consistent set is a deductively closed, consistent set that contains, for every formula A, either A or $\sim A$ $(A \supset \bot)$.

The left-right direction follows immediately from the definition of R.

To prove the right-left direction let $M' = \{m' : A \in m' \ \&\ mRm'\}$, that is M' contains all worlds accessible from m at which A is true. Suppose further that $B \in m'$ for every $m' \in M'$. The $m' \in M'$ are all maximally consistent sets containing $\{C : A \to C \in m\} \cup \{A\}$, and since B is in every m' it follows that $\{C : A \to C \in m\} \cup \{A\} \vdash B$. But then $\{A \to C : A \to C \in m\} \cup \{A \to A\} \vdash A \to B$ (as shown above). $A \to A$ is a theorem of the logic and $\{A \to C : A \to C \in m\} \subseteq m$, therefore $A \to B \in m$ by the deductive closure of m.

The two deductions

$$\cfrac{\cfrac{}{\top} \qquad \cfrac{\cancel{\top}^1 \quad \cancel{\top \to A}^1}{A}\ {\to}E}{(\top \to A) \to A}\ {\to}I(1) \qquad\qquad \cfrac{\cfrac{\cancel{\top \to A}^2 \quad \cancel{\top \to A}^1}{\top \to (\top \to A)}\ {\to}I(1)}{(\top \to A) \to [\top \to (\top \to A)]}\ {\to}I(2)$$

show that $(\top \to A) \to A$ and $(\top \to A) \to [\top \to (\top \to A)]$ are theorems. So $\{A : \top \to A \in m\} \subseteq m$ thence mRm. Also if, mRm' and $m'Rm''$ then if $\top \to A \in m$ then $\top \to (\top \to A) \in m$ so $\top \to A \in m'$ so $A \in m''$, hence $\{A : \top \to A \in m\} \subseteq m''$ so mRm''. Taking all deductively closed sets as the domain we can therefore construct a Kripke model for the logic.

The remaining connectives are easy, as each m is consistent, deductively closed and must contain either A or $\sim A$ for any A. $A \wedge B \in m$ iff $A \in m$ and $B \in m$. $A \vee B \in m$ iff $A \in m$ or $B \in m$ (as if neither is in m then $\sim A, \sim B \in m$ and m is inconsistent). $A \supset B \in m$ iff $\sim A \in m$ or $B \in m$.

Treating truth as membership we may use each m as a model for $\mathcal{L}$ (with sentences of the form $B \to C$ treated like new atomic formulae).[11] Hence we obtain a model of the desired form.

So if in $\mathcal{L}_i$, $\Gamma \nvdash A$ then there is a deductively closed set containing Γ and not A, and so $\Gamma \cup \{\sim A\}$ is consistent (if it is inconsistent then we may use *restart* or PIP to deduce A from Γ). So $\Gamma \cup \{\sim A\}$ can be extended to a maximal consistent set. This set will appear somewhere in the canonical model above and so A is false in some model in which Γ is true. ∎

We may now define strict negation $\neg$ so that $\neg A$ is equivalent to (or a shorthand for) $A \to \bot$. The system is as close as I can come to adding an

[11] Since the maximally consistent sets contain a negation satisfying the classical rules they will do for models of the classical connectives.

intuitionistic conditional to classical logic.[12] We would be using the term somewhat loosely to call $\to$ of $\mathcal{L}_i$ *intuitionistic implication* (and $\neg$ intuitionistic negation), as not every theorem of intuitionistic logic is a theorem of $\mathcal{L}_i$. For example

$$A \to (B \to A)$$

is not a theorem and the inference

$$\frac{A}{B \to A}$$

is not valid. If A is strong (e.g. A itself is an implication) then both are theorems/valid. If we set all atomic formulae as strong then we obtain a logic closer to intuitionistic logic as $A \to B \to A$ is a theorem when A is atomic. But I see no reason to do this.

We may treat $\Box$ as shorthand for $\top \to A$ then we will obtain the familiar language for the modal logic $S4$.

10.2.4 Deduction theorem

The deduction theorem does not hold in $\mathcal{L}_i$. If A is weak then although $A, B \vdash A$, it is not the case that $A \vdash B \to A$. The deduction theorem is retained for the material conditional $\supset$ however.

10.3 Normalisation

An introduction and subsequent elimination of $\to$ may be removed in the normal way:

$$\dfrac{\begin{array}{c}\vdots\\ A\end{array} \qquad \dfrac{\begin{array}{c}\vdots\\ C_1 \ldots C_n\end{array} \qquad \begin{array}{c}\cancel{A}^m, \cancel{C_1}^m \ldots \cancel{C_n}^m\\ \vdots\\ B\end{array}}{A \to B}\ \to I(m)}{B}\ \to E$$

gets reduced to

$$\begin{array}{c}\vdots \qquad \vdots\\ A, C_1 \ldots C_n\\ \vdots\\ B\end{array}$$

[12] Interestingly, we get a slightly different logic if we add classical negation to intuitionistic logic. The difference is in the interpretation of atomic formulae. If we begin with intuitionistic logic (and add a classical conditional) then the logic we get is the same as $\mathcal{L}_i$ except that all atomic formulae are strong.

however there is another way a connective could be eliminated for the introduction rule $\to I$ serves to eliminate some of its minor premises (the C_i):

$$\dfrac{\dfrac{\vdots}{C_1}\,?I \quad \ldots C_n \qquad \begin{matrix}\cancel{A}^m, \cancel{C_1}^m \ldots \cancel{C_n}^m \\ \vdots \\ B\end{matrix}}{A \to B} \to I(m)$$

and C_1 has been introduced by $?I$ and then eliminated. However, the condition that all the C_i must be strong means that it is of the form $D_1 \to D_2$ and so $?I$ is the rule $\to I$:

$$\dfrac{\dfrac{\begin{matrix}\vdots \\ E_1 \ldots E_m\end{matrix} \qquad \begin{matrix}\cancel{D_1}^k, \cancel{E_1}^k \ldots \cancel{E_m}^k \\ \vdots \\ D_2\end{matrix}}{D_1 \to D_2} \to I(k) \quad \ldots C_n \qquad \begin{matrix}\cancel{A}^m, \cancel{C_1}^m \ldots \cancel{C_n}^m \\ \vdots \\ B\end{matrix}}{A \to B} \to I(m)$$

and, since the E_i must also be strong, there is a deduction of C_1 (i.e. $D_1 \to D_2$) from the E_i, we can use it to produce the following deduction:

$$\dfrac{\begin{matrix}\vdots \\ E_1 \ldots E_m\end{matrix} \quad \begin{matrix}\vdots \\ C_2 \ldots C_n\end{matrix} \qquad \begin{matrix} & \cancel{D_1}^m, \cancel{E_1}^m \ldots \cancel{E_m}^m & \\ & \vdots & \\ \cancel{A}^m, & D_1 \to D_2 & , \cancel{C_2}^m \ldots \cancel{C_n}^m \\ & \vdots & \\ & B & \end{matrix}}{A \to B} \to I(m)$$

10.4 The paradoxes of material implication and the strict conditional

There are problems with the view that classical logic presents the full story of the logical connectives. The greatest problem lies in the classical theory of implication: $\supset$. Material implication has too strong a logic to be what is meant by natural language 'if…then…'. It is natural then to think that the natural language conditional is in fact a strict conditional (or sometimes strict and sometimes material). Unfortunately this view is also problematic and I conclude (in chapter 13) that the logic of the natural language conditional is much more subtle.

Nevertheless the view that the natural language conditional is a strict conditional is not untenable and has been defended well in the literature (e.g. Lowe).

It is interesting to see how $\mathcal{L}_i$ is free from familiar objections to analysing 'if... then... ' as $\supset$.

10.4.1 Some deductions with strict conditionals

Since we cannot assume that A or $\sim A$ are strong in general,

$A \nvdash_{\mathcal{L}_i} B \rightarrow A$
${\sim}A \nvdash_{\mathcal{L}_i} A \rightarrow B$

However, if A is strong then

$A \vdash_{\mathcal{L}_i} B \rightarrow A$

which is not unacceptable, for example $\ulcorner$if B then $1+1=2\urcorner$ is true for any B, or at least for any B that is possible. Perhaps a little more contentiously, if A is impossible then if it is true, anything is, i.e.

$\neg A \vdash_{\mathcal{L}_i} A \rightarrow B.$

It may be thought that the natural language conditional is false in the case of an impossible antecedent, or at least not assertable. I think this is not in general true, for otherwise a conditional like 'if $1+1=3$ then a contradiction follows' is false, which is not the case such conditionals are true. However, the conditional that requires a *possible* antecedent may be characterised by

(‡) $({\sim}\neg A) \wedge (A \rightarrow B)$.

10.4.2 Other deductions with strict conditionals

Clearly $A \rightarrow B$ entails $A \supset B$, but not vice versa.

Disjunctive syllogism is valid in $\mathcal{L}_i$,

$$\dfrac{A \vee B \qquad \dfrac{\dfrac{\cancel{A}^1 \quad {\sim}A}{\bot}\,{\sim}E}{B}\,\bot E \qquad \cancel{B}^1}{B}\,\vee E(1)$$

but we cannot introduce ${\sim}A \rightarrow B$ from this as the premise $A \vee B$ might not be strong (and does not have a strong formula below it that does not depend on ${\sim}A$). So although

$A \vee B, {\sim}A \vdash_{\mathcal{L}_i} B$, and even $A \vee B, \neg A \vdash_{\mathcal{L}_i} B$

the restriction is not in general met to allow the conditional proof for the strict conditional:

$$A \vee B \nvdash_{\mathcal{L}_i} \sim A \rightarrow B, \text{ and also } A \vee B \nvdash_{\mathcal{L}_i} \neg A \rightarrow B.$$

If $A \vee B$ is strong then the restriction is met and we may introduce the intuitionistic conditional in the usual way. Since there is no restriction on the classical conditional

$$A \vee B \vdash_{\mathcal{L}_i} \sim A \supset B, \text{ and also } A \vee B \vdash_{\mathcal{L}_i} \neg A \supset B.$$

In $\mathcal{L}_i$, $A \rightarrow B$ is equivalent to $\neg(A \wedge \sim B)$, but not to $\sim(A \wedge \sim B)$ as

$$\sim(A \wedge \sim B) \nvdash_{\mathcal{L}_i} A \rightarrow B$$

and similarly

$$\sim A \vee B \nvdash_{\mathcal{L}_i} A \rightarrow B$$

although the classical (material) conditional $A \supset B$ is entailed by $\sim(A \rightarrow \sim B)$ and $\sim A \vee B$.

10.5 Conclusion

I have shown here that even operating within the constraint of harmony we can give an account of the strict conditional. We are not confined to the basic classical logic by demanding normalisation.

I shall not argue that the only conditionals that occur in natural language are the material conditional and the strict conditional, but I hope to have shown that an account along those lines (e.g. and account with a variety of strict conditionals) looks like it could provide a good analysis of the conditional used in natural language. In particular, we can account for the analyticity of much reasoning with the strict conditional (as we can give harmonious rules for it).

Chapter 11

First order logic

So far I have considered mainly propositional logic, in this chapter I shall discuss issues sounding the familiar rules for the quantifiers. I shall present a normalised system for first order logic which will allow me to complete my argument that the consequences of classical first order logic are analytic.

Actually I shall not complete my argument there, for I think that a discussion of first order logic is not complete without some comment on the phenomenon of reference failure. I shall suggest a logic that allows us to handle empty proper names, I propose something similar to familiar *negative free logics* (see [**?**]). Strictly speaking, therefore, I do not conclude that classical logic is analytic for classical logic has no empty reference and I present a logic that allows for it. But the logic I present is similar enough, I think, to classical logic to warrant falling under the title of this thesis.[1]

11.1 First order rules

11.1.1 Universal quantification

A point of difficulty arises in using the natural deduction system in the generalisation to first order logic. The familiar rules for the universal quantifier are complicated slightly by the addition of restart:

$$\frac{\forall x A}{A[x/t]}\ \forall E \quad \text{provided } t \text{ is free for } x \text{ in } A \qquad \frac{A}{\forall x A[y/x]}\ \forall I \quad \text{provided } y \text{ is not free in any formulae or assumptive rule applications on which } A \text{ depends}$$

[1]For example, the logic of reference failure I present is not strictly speaking a free logic as it does entail $\exists x(x = x)$. Usually, only logics that can have an empty domain (and not entail $\exists x(x = x)$) are called ‘free logics’.

y is *free for x in A* when no free occurrence of x in A in the scope of a quantifier binding y. That is, any free occurrence of x in A may be replaced by y which remains free. A term t is free for x in A when all its free variables are free for x in A.[2] The side-condition on $\forall I$ ensures that y is not free in the premises of any incomplete applications of restart on which A depends.[3]

We may extend the justification of the logical connectives to these rules. A speaker may think that A, and not intend to make reference to any object in particular when asserting A. For example, a speaker may be believe that something has spots only on learning that it is a cow, without waiting to find out which cow it is. The verbalisation of '... but it does not matter what object x is' is $\forall x$. Similarly if the speaker believes arbitrarily that a cow has spots then he is being true to his belief to believe of a particular cow that it has spots, and true to his intentions if he infers it from his belief that *all cows* have spots.[4] Certainly, when described the rules are obvious. If everyone had a parent, then so have I. Now given any person, nobody in particular, they have a parent. A simpler way of putting the last sentence is 'everyone has a parent'. Actually, serves better as a justification of a generalised quantifier meaning 'all... are...': let $\forall x.A.B$ be a binary quantification, then these are the rules for it (I omit the rule name and superscripts):

$$\frac{\forall x.A.B \quad A[x/t]}{B[x/t]} \qquad \frac{\begin{array}{c}\cancel{A}\\ \vdots \\ B\end{array}}{\forall x.A[y/x]x.B[y/x]}$$

provided B does not depend on any assumptions or weak rule applications in which y is free, except A

It is easy to show that these rules entail that $\forall x.A.B$ is equivalent to $\forall x(A \supset B)$. For simplicity I shall base the remaining discussion around the unary quantifier $\forall x$.

[2] The restrictions about t being free for x are to ensure that we do not add or bound variables with our replacements. For example from $\forall x \exists y Rxy$ we cannot apply $\forall E$ setting t to be y, for then we conclude $\exists y Ryy$ which does not follow from $\forall x \exists y Rxy$. Notice that y is not free for x in $\exists y Rxy$.

[3] I formulate the rule in terms of assumptive rule applications so that we would not have to modify it were we to add any rules other than restart that have assumptive applications. I shall not add any such rules in this thesis, but I think it is a virtue to maintain a certain degree of generality in presenting these rules.

[4] For the term 'all cows... ' is introduced with the intention of verbalising that what follows (what goes in place of '... ') does not depend on which cows it refers to.

11.1.2 Universal harmony

Harmony in the case of the universal quantifier is easy enough. If the connective is introduced and then eliminated thusly:

$$\dfrac{\dfrac{\begin{matrix}\vdots\\ A\end{matrix}}{\forall x A[y/x]}\ \forall I}{A[y/t]}\ \forall E$$

then the side-condition on the introduction rule ensures that, after suitable renaming bound variables, we may replace y by t in the Prawitz tree as far as A allowing us to delete the introduction and subsequent elimination of $\forall x$:

$$\begin{matrix}\vdots\\ A[y/t]\end{matrix}$$

When y is replaced by t throughout the deduction of A we may have to change some bound variables (so that nothing in t gets bound when we put it in place of y).

11.1.3 Replacement of variables

Normalisation theorems are complicated by the universal quantifier. The problem lies in the reduction cases for $\rightarrow$, $\supset$ and $\vee$ and $\exists$ (below).

Call a variable x *critical* in a Prawitz tree if it occurs free in the Prawitz tree is not free in any of the formulae at top nodes of the tree that are not crossed out.

Now, if we introduce and then eliminate $\supset$:

$$\dfrac{\begin{matrix}\vdots\\ A\end{matrix} \qquad \dfrac{\begin{matrix}\cancel{A}^n\\ \vdots\\ B\end{matrix}}{A \supset B}\ \supset I(n)}{B}\ \supset E$$

then before adding the deduction of B from A (the premise of $\supset I$) on the end of the deduction of A (the minor premise of $\supset E$), we must replace some variables in the deduction of B from A. We must replace uniformly all the variables $x_1 \ldots x_n$ that are critical in the deduction of B from A by $y_1 \ldots y_n$, where no y_i occurs in the deduction of A. After making these replacements we can reduce the introduction and subsequent elimination to:

$$\begin{array}{c}\vdots\\ A\\ \vdots\\ B\end{array}$$

and any application of $\forall I$ in the deduction from A to B is sure to have its side condition met.

Similar considerations apply to the reduction cases for $\rightarrow$, $\supset$ and $\vee$ and $\exists$ below.

11.1.4 Existential Quantification

The rules for the existential quantifier are more problematic:

$$\frac{\exists x A \qquad \begin{array}{c}\cancel{A[x/c]}^{n}\\ \vdots\\ B\end{array}}{B}\ \exists E(n)$$

where c is a constant that does not occur in any formulae or assumptive rule applications on which B depends (except $A[x/c]$), nor in $\exists x A$ nor in B itself.

$$\frac{A[x/t]}{\exists x A}\ \exists I$$

provided t is free for x in A

The idea behind the restriction on the elimination rule is to make it a generalisation of this argument

$$\frac{\begin{array}{l}\text{Something is } A\\ \text{Anything that is } A \text{ is in a world where } B \text{ is true}\end{array}}{B}$$

thus c is used as an arbitrary object that satisfies A so as to capture the ‘anything’ without using a universal quantifier. In order to ensure that c is arbitrary it must not already appear in A. Neither must c appear in any formulae on which B depends, this is similar to the restriction on the introduction rule of the universal quantifier.

The rule $\exists E$ is problematic because it seems hard for anybody to understand let alone apply correctly. Furthermore, listening to people reasoning existential statements, it is hard to spot how where this rule is used explicitly. I think the rule is used explicitly but in a more roundabout way than the other rules.

Suppose I return home one evening and notice a light on in my house and sense movement inside. I might immediately jump to the conclusion that:

There is someone in my house

when reasoning further from this I typically use pronouns like 'he' or 'it' or descriptions like 'the person in my house', for example I might conclude:

All my family are abroad so *it* is not a relation of mine.
Nobody else has a key so *the person in my house* broke in.
I should call the police.

notice that I can sensibly use a directly referring pronoun 'it' or 'he' without a determinate referential link to anything (e.g. if there are two burglars in my house, to whom does 'it' refer?), and I sensibly use the definite article 'the' without uttering a definite description (again, there may be two burglars). These deviant uses of 'he' and 'the' are sensible in the context of me reasoning from an existential belief that someone is in my house. Only the sentence 'I should call the police' is fully intelligible without knowledge of this context (although it may not be clear why the police should be called). My use of the terms 'it', 'he' and 'the person in my house' are more to facilitate reasoning than as fully fledged referring expressions. I should not be worried about my use of the phrase 'the person in my house' when I discover there were really two burglars because I never intended it as a genuine definite description, I was using it to aid my reasoning (which concluded in 'I should call the police').

But this is exactly what the existential elimination rule does. To eliminate $\exists x A$ we assume, for a side deduction (i.e. in the context of the existential statement $\exists x A$), that $A[x/c]$. That is, we find a neutral referring expression (neutral as in we assume nothing about it except that it satisfies A) and say that $A[x/c]$. Since c is meaningful only in the context of facilitating reasoning from $\exists x A$ we can end our deduction only when we have stopped using c. To ensure neutrality of c we must know nothing about it except that it is A, more formally it must not be free in $\exists x A$ nor in any *other* assumptions we use to deduce B from $A[x/c]$. Finally to ensure we conclude with something that has truth conditions we must ensure that c is not (free) in B.

The simplest method of finding a neutral term c that we can keep track of (to make sure we have made no extra assumptions about it) is to use a pronoun or a description like 'the A', although such a uses are not really referring expressions. And, of course, people are lazy and very often to not bother freeing their assertions from the context. I may end my reasoning from 'someone is in my house' with the conclusion 'he is a burglar', it is clear from the context that the inference is to be concluded silently with 'there is someone in my house who is a burglar'. It is far easier to think 'there is someone in my house who is a burglar' than it is to say, saying it takes up more time and energy which is perhaps why it is hard to find spoken instances of the existential weaken rule, but for me a silent use is good enough. We might extend the rule $\exists E$ to this:

$$\dfrac{\exists x A \qquad \begin{array}{c}\cancel{A[x/c]}^{n}\\ \vdots \\ B\end{array}}{B}\ \exists E(n)$$

where c is a constant that does not occur in any formulae on which B depends, except $A[x/c]$ *and any assumption anaphoric on* $\exists x A$, and c does not occur in $\exists x A$ nor in B itself.[5]

Consider the sentences

There is a fly in my soup. It is still alive.

the first of these is in existential quantification $\exists x[\text{fly}(x) \wedge \text{in my soup}(x)]$, and the second sentence is *anaphoric* on the first. We might analyse them as two separate assumptions:

$$(1) \quad \exists x[\text{fly}(x) \wedge \text{in my soup}(x)]$$

$$(2) \quad \text{still alive}(c)$$

but add that the second is anaphoric on the first, so that c may be used together with (2) when eliminating (1):

$$\dfrac{\exists x[\text{fly}(x) \wedge \text{in my soup}(x)] \qquad \dfrac{\dfrac{\cancel{\text{fly}(c) \wedge \text{in my soup}(c)}^{1} \qquad \text{still alive}(c)}{\text{fly}(c) \wedge \text{in my soup}(c) \wedge \text{still alive}(c)}\ \wedge I}{\exists x[\text{fly}(x) \wedge \text{in my soup}(x) \wedge \text{still alive}(x)]}\ \exists I}{\exists x[\text{fly}(x) \wedge \text{in my soup}(x) \wedge \text{still alive}(x)]}\ \exists E(1)$$

which allows us to conclude, as we should, that

There is a live fly in my soup

from

There is a fly in my soup. It is still alive.

[5] This side-condition is, as things stand, not legitimate. This is because no effective method has been given of determining what is anaphoric on what just by looking at a Prawitz tree. A simple and unsubtle way of overcoming this difficulty is to add to each Prawitz deduction a list of exactly what is anaphoric on what. A more subtle and complicated way of handling anaphora is to use the epsilon calculus.

11.1.5 Existential harmony

Harmony for the existential quantifier is unproblematic:

$$\dfrac{\dfrac{\begin{array}{c}\vdots\\ A[x/t]\end{array}}{\exists x A}\,\exists I \qquad \begin{array}{c}\cancel{A[x/c]}^n\\ \vdots\\ B\end{array}}{B}\,\exists E(n)$$

may be reduced by replacing c by t and, after suitable replacement of critical variables (see section 11.1.3), appending the deduction of B from $A[x/t]$ to the deduction of $A[x/t]$.

$$\begin{array}{c}\vdots\\ A[x/t]\\ \vdots\\ C\end{array}$$

The restrictions on our choice of c in $\exists E$ ensures that it may be replaced by t without affecting the validity of the deduction. We may also define a binary existential quantifier:

$$\dfrac{\exists x.A.B \qquad \begin{array}{c}\cancel{A[x/c]}, \cancel{B[x/c]}\\ \vdots\\ C\end{array}}{C}$$

where c is a constant that does not occur in any formulae or assumptive rule applications on which C depends (except $A[x/c]$ and $B[x/c]$), nor in $\exists x.A.B$ nor in C itself.

$$\dfrac{A[x/t] \quad B[x/t]}{\exists x.A.B}$$

provided t is free for x in A and B

11.2 Identity

The rules for identity are these:

$$\dfrac{}{t = t}\,{=}I \qquad \dfrac{t_1 = t_2 \quad A[x/t_1]}{A[x/t_2]}\,{=}E$$

provided t_1 and t_2 are free for x in A

we may also add an extra elimination rule, but only shortens some deductions without adding any deductive power:

$$\dfrac{t_1 = t_2 \quad A[x/t_2]}{A[x/t_1]}\,{=}E$$

provided t_1 and t_2 are free for x in A

Harmony is easy to show, if $t_1 = t_2$ is introduced and then eliminated then t_1 and t_2 are syntactically identical:

$$\dfrac{\overline{t = t} \quad \begin{array}{c}\vdots \\ A[x/t]\end{array}}{A[x/t]}\, {=}E$$

and we may replace the whole thing with the deduction of $A[x/t]$:

$$\begin{array}{c}\vdots \\ A[x/t]\end{array}$$

11.3 Reference failure

It would be well objected that $t = t$ is not analytic as it is not true when t does not refer. I am sympathetic to the view that ‘John is John’ or ‘John swims’ are not true if John does not refer.

A simple response to this is to deny that ‘John is John’ expresses a proposition if ‘John’ does not refer, with this we can avoid the objection by arguing that we are interested in accounting for the analyticity of sentences that express propositions.[6] I think this a reasonable thing to say about the apparent failure of the law of excluded middle in the case of vague predications. But I think in the case of reference failure it is too simple a response, at least with respect to questions about the relationship between analyticity and modality. In the case of a vague predication, if it is vague that a is F in one possible situation then (it is commonly supposed) it is vague that a is F in all other possible situations (see [**?**]). However if a does not exist in one possible situation it might still exist in others. This suggests that even if $\ulcorner a \urcorner$ does not refer the sentence $\ulcorner a$ exits$\urcorner$ does express a proposition. Also consider a sentence like ‘Napoleon no longer exists’ which seems to express a true proposition.

I shall suppose here that such sentences do express propositions and we should account for non-referring terms in the theory of analyticity.[7] I shall now present a modification to classical logic that allows us to account for reference failure.

[6]For example ‘splurg or not splurg’ is not analytic because it expresses no proposition at all.

[7]The matter is complicated further by some philosophers suggesting that ‘Napoleon’ does refer, even though Napoleon is dead (and presumably no longer exists). I shall not discuss these issues here.

11.3.1 A simple analysis

The simplest way of allowing for non-referring constants is to modify the quantifier and identity rules to obtain a free logic. Technically what I shall now produce is not a free logic, although it can handle empty proper names, because $\exists x(x = x)$ is a theorem. Free logics are easy enough to produce, more interesting is a logic that allows reference failure and allows the *possibility* of an empty domain, but recognises that *this* domain is not empty.[8]

I shall take it as given that an atomic formula $F(t_1 \ldots t_n)$ is not true

---

[8]Tennant, in [**?**, p167-175] and in other of his arguments and papers, presents a free logic. I shall make some points of comparison here of his logic and mine, my logic is just like his except with some additional subtleties.

1. My free logic is not really a free logic as $\exists x(x = x)$ is a theorem, in Tennant's logic it is not a theorem. My logic has the additional subtlety that $\Box\exists x(x = x)$ is not a theorem and $\Diamond\sim\exists x(x = x)$ is consistent. So although it is analytic that something exists (I think that 'I exist' is analytically true) the logic allows *the possibility* that nothing exists. Tennant does not present a modal free logic.

2. The logic I present is *negative*, that is, if a has no reference then Fa is false if F is an atomic predicate. Tennant's logic is not negative, although it is easy enough to make it so.

3. Following on from the first point, in order to say that t refers in Tennant's system we write $\exists x(x = t)$ whereas in my system we need only write $t = t$. Furthermore, if t is a variable then it refers ($\exists x(x = y)$ is a theorem of my free logic, but again, $\Box\exists x(x = y)$ is not). This makes for some more elegant quantifier rules, for example here is Tennant's $\forall I$ rule:

$$\begin{array}{c} \cancel{\exists x(x = y)} \\ \vdots \\ \dfrac{A[x/y]}{\forall x A} \end{array}$$

with the side condition that y not occur free in any other assumptions on which $A[x/y]$ depends. My rule for the universal quantifier has no such discharging. This means that if the rules for the universal quantifier constitute its definition then it is not defined in terms of the existential quantifier. Also the introduction rule for the existential quantifier does not seem so circular, here is Tennant's:

$$\frac{\exists x(x = t) \quad A[x/t]}{\exists x A}$$

My introduction rule is more elegant because the existential quantifier appears only in the conclusion. These inelegancies in Tennant's logic become highly objectionable in the context of the theory of implicit definitions I propose in chapter 14 (and indeed in the accounts given by Dummett and Hacking) where I argue that such definitions must be constituted by introduction and elimination rules in a strict sense (e.g. no circularity of the type displayed by Tennant's rules is allowed).

unless the terms $t_1 \dots t_n$ all refer.[9]

The rules for $\forall$ become:

$$\frac{\forall x A \quad B}{A[x/t]}\ \forall E$$

provided t is free for x in A, and B is an atomic formula in which t is free. If t is a variable then B need not be present.

$$\frac{A}{\forall x A[y/x]}\ \forall I$$

provided y is a variable and A does not depend on any assumptions or assumptive rule applications in which y is free

The rules for $\exists$ become:

$$\frac{\exists x A \qquad \begin{matrix} \cancel{A[x/y]}^n, \\ \vdots \\ B \end{matrix}}{B}\ \exists E(n)$$

where y is a variable that does not occur free in any formulae or assumptive rule applications on which B depends (except $A[x/c]$), nor free in $\exists x A$ nor free in B itself.

$$\frac{A[x/t] \quad B}{\exists x A}\ \exists I$$

provided t is free for x in A, and B is an atomic formula in which t is free. If t is a variable then B need not be present.

The rules for identity become:

$$\frac{B}{t = t}\ {=}I$$

where B is an atomic formula in which t is free. If t is a variable then B need not be present.

$$\frac{t_1 = t_2 \quad A[x/t_1]}{A[x/t_2]}\ {=}E$$

provided t_1 and t_2 are free for x in A

So for example:

$$\frac{}{x = x}\ {=}I$$

is a deduction as x is a variable, but if x is replaced by a constant the deduction is not valid.

11.3.2 Normalisation

Harmony follows pretty much as before. If $\forall$ is introduced and then eliminated

$$\frac{\dfrac{A}{\forall x A[y/x]}\ \forall I \qquad \begin{matrix} \vdots \\ B \end{matrix}}{A[y/t]}\ \forall E$$

then we may replace y by t in the inference of A. But doing so may invalidate some applications of $\forall E$, $\exists I$ and $=I$ that relied on y being a variable. We

[9] As does Burge in [?].

can revalidate them by adding B as a premise, so the inference reduces to:

$$\begin{array}{c}\vdots\\ B\\ \vdots\\ A[y/t]\end{array}$$

where instances of

$$\frac{}{y=y}=I \qquad \frac{\forall xC}{C[x/y]}\,\forall E \qquad \frac{D}{\exists xD[y/x]}\,\exists I$$

have been replaced by

$$\frac{B}{t=t}=I \qquad \frac{\forall xC \quad B}{C[x/t]}\,\forall E \qquad \frac{D[x/t] \quad B}{\exists xD[t/x]}\,\exists I$$

If $\exists$ is introduced and then eliminated

$$\cfrac{\cfrac{\begin{array}{c}\vdots\\ A[x/t]\end{array} \quad \begin{array}{c}\vdots\\ B\end{array}}{\exists xA}\,\exists I \qquad \begin{array}{c}\cancel{A[x/y]}^n\\ \vdots\\ C\end{array}}{C}\,\exists E(n)$$

then we may replace y by t in the inference of C and, as with the universal quantifier, we must add B as a minor premise to any rule that relied on y being a variable.

$$\begin{array}{c}\begin{array}{cc}\vdots & \vdots\\ A[x/t] & B\end{array}\\ \vdots\\ C\end{array}$$

Finally if $t=t$ is introduced, it can be eliminated by $=E$:

$$\cfrac{\cfrac{B}{t=t}=I \quad A[x/t]}{A[x/t]}=E$$

the reductions for this is easy. Also $t=t$ could be eliminated by an application of $\forall E$, $\exists I$ and $=I$:

$$\cfrac{\cfrac{B}{t=t}=I}{t=t}=I \qquad \cfrac{\forall xA \quad \cfrac{B}{t=t}=I}{A[x/t]}\,\forall E \qquad \cfrac{A[x/t] \quad \cfrac{B}{t=t}=I}{\exists xA}\,\exists I$$

the reductions for these are equally easy, just delete the introduction of $t=t$:

$$\frac{B}{t=t}=I \qquad \frac{\forall xA \quad B}{A[x/t]}\,\forall E \qquad \frac{A[x/t] \quad B}{\exists xA}\,\exists I$$

11.3.3 The result

The logic we get from this is close to a standard negative free logic where every atomic proposition involving a non-referring term is false.[10] For example:

$$\dfrac{\dfrac{\sim\exists x(x=t) \qquad \dfrac{\dfrac{\cancel{Ft}^1}{t=t}\,{=}I \quad \cancel{Ft}^1}{\exists x(x=t)}\,\exists I}{\bot}\,\supset E}{\sim Ft}\,\sim I(1)$$

Using normalisation we can show that $\exists x(x = t)$ is not a theorem where t is not a variable. A normalised deduction of it must finish with by application of $\exists I$ to $t' = t$, if t is not a variable there is no way to deduce anything of the form $t' = t$ from no assumptions.

11.3.4 Strict implication

In order to accommodate strict implication appropriately, we must add that some applications of $\forall E$, $\exists I$ or $=I$ are weak rule applications.

> Any application of $\forall E$, $\exists I$ or $=I$ in which the term introduced in the conclusion does not appear in the atomic premise (if there even is such a premise) because it is a variable, is a weak rule application.

For example:

$$\dfrac{\forall x A}{A[x/y]}\,\forall E$$

is a weak rule application.

$a = b$ does not entail $\Box(a = b)$, for we cannot deduce $\Box(a = a)$. This is because $a = a$ is not a theorem. Also, we can use $=I$ to deduce $x = x$ but since this would be a weak rule application we cannot use it to deduce $\Box(x = x)$. $a = b$ does entail that $a = a \rightarrow a = b$ by applying $=E$ to a deduction of $a = a \rightarrow a = a$. That is, from $a = b$ we cannot deduce that a is necessarily b, but we can deduce that necessarily if a exists then it is b.

---

[10] See Burge's paper [**?**] for such a logic, a difference between his and mine is that $\exists x(x = x)$ is a theorem of my logic and not of Burge's. From the way I have set things up, although $\exists x(x = x)$ is a theorem, $\Box\exists x(x = x)$ is not. This means that the logic acknowledges that the universe *could* be empty but *actually* is not.

11.3.5 Semantics

A model consists of a domain D, a set of subsets (call them worlds) of the domain W, a reflexive and transitive accessibility relation R on the worlds, and an interpretation I of the constants, predicates and function symbols of the language.

I assigns to each n-ary function symbol a function from D^n to D and to each n-ary predicate symbol F a set of $n+1$-tuples of the form $\langle w, e_1 \ldots, e_n \rangle$ where w is a world in W and each $e_i \in w$. That is I fixes the extension of F at each world.

For simplicity we shall assume that the language has only constants and no function symbols (except for the constants). One world $w_a \in W$ is the actual world.

A valuation v^w on a model M assigns elements of the world w to every variable.

Let $M \vDash_w^v A$ mean that A is true at world w in model M under valuation v. Let x^v denote the object assigned by v to the variable x. If t is a constant then t^I is the object I assigns to t, and $t^v = t^I$. Also let F^I be the set I assigns to F.

The semantics are little different from the usual semantics for first order modal logic.

$M \vDash_w^v Ft_1 \ldots t_n$ iff $\langle t_1^v \ldots t_n^v \rangle \in F^I$

$M \vDash_w^v t_1 = t_2$ iff $t_1^v = t_2^v$

$M \vDash_w^v A \wedge B$ iff $M \vDash_w^v A$ and $M \vDash_w^v B$

$M \vDash_w^v A \vee B$ iff $M \vDash_w^v A$ or $M \vDash_w^v B$

$M \vDash_w^v A \supset B$ iff either $M \nvDash_w^v A$ or $M \vDash_w^v B$

$M \nvDash_w^v \bot$

$M \vDash_w^v \forall x A$ iff $M \vDash_w^{v'} A$ for every valuation v' that is x-alternate to v and that assigns an element of w to x.[11]

$M \vDash_w^v \exists x A$ iff $M \vDash_w^{v'} A$ for some valuation v' that is x-alternate to v and that assigns an element of w to x.

$M \vDash_w^v A \rightarrow B$ iff $M \vDash_{w'}^v B$ for every w' s.t. wRw' and $M \vDash_w^v A$.

[11] v' is x-alternate to v when it agrees with v on what is assigned to every variable except perhaps x.

> $M \vDash A$ iff $M \vDash^v_{w_a} A$ for every valuation v that assigns only elements of w_a (the actual world) to the variables.

Note that $M \vDash^v_w A \to B$ iff $M \vDash^v_{w'} A \to B$ for all w' s.t. wRw', so if a strong formula is true at w it is true at all future w'.

Soundness

It is easy to see that the rules are sound for this semantics.[12]

First we must show that if $M \vDash^v_w \Gamma$ and Γ entails A without any variable dependent rule applications then $M \vDash^v_w A$ for any w. The proof of this is by induction on the length of the deduction.

For example, suppose the deduction ends with an application of $\to I$:

$$\dfrac{C_1 \ldots C_n \qquad \begin{matrix} \cancel{A}, \cancel{C_1} \ldots \cancel{C_n} \\ \vdots \\ B \end{matrix}}{A \to B} \to I$$

by induction hypothesis $M \vDash^v_w C_i$ and if $M \vDash^v_{w'} \{A, C_1 \ldots, C_n\}$ then $M \vDash^v_{w'} B$ for every w', since each C_i is strong $M \vDash^v_{w''} C_i$ for every w'' s.t. wRw''. And so $M \vDash^v_{w'} B$ for every w' s.t. wRw' and $M \vDash^v_{w'} A$.

If the deduction ends by an application of $\exists I$:

$$\dfrac{A[x/t] \quad B}{\exists x A} \exists I$$

then by induction hypothesis $M \vDash^v_w A$ and $M \vDash^v_w B$, since B is atomic and t is free in B (we are assuming that $\exists I$ is not variable dependent), $t^v \in w$, and so there is a valuation v' x-alternate to v which assigns an element of w (t^v) to x.

The other cases are shown similarly.

Now we can show that if $M \vDash^v_{w_a} \Gamma$ and Γ entails A then $M \vDash^v_{w_a} A$, where v assigns only elements of w_a to the variables. Once again this is shown by induction of the length of the deduction.

The interesting cases are variable dependent rule applications and $\to I$. For example

[12]The argument I present here is for a logic where *restart* is replaced by PIP. See section 8.2 to see how *restart* and PIP are interchangeable (i.e. switching one for the other does not alter the logical consequence relation). In the following proof I omit the clause for PIP.

$$\overline{x = x}$$

is sound as, by assumption, every variable is assigned a member of w_a.

And if the deduction ends with an application of $\rightarrow I$, by induction hypothesis $M \vDash^v_{w_a} C_i$ and if $M \vDash^v_{w'} \{A, C_1 \ldots, C_n\}$ then $M \vDash^v_{w'} B$ for every w' (we can assume this as the deductive premise makes use of no variable dependent rules), since each C_i is strong $M \vDash^v_{w''} C_i$ for every w'' s.t. $w_a R w''$. And so $M \vDash^v_{w'} B$ for every w' s.t. $w_a R w'$ and $M \vDash^v_{w'} A$.

The other cases are also easily verified.

Completeness

Say that $\Gamma \vdash A$ when Γ deduces A in the free logic above, say $\Gamma \vdash_s A$ when Γ deduces A without any variable dependent rule applications.

First we must extend the language to ensure that there is a different constant c for every formula $\exists x A$ of the language.

We take all sets m such that either A or $\sim A$ is in m and m is consistent with respect to $\vdash_s$, and such that if $\exists x A \in m$ then $A[x/c], c = c \in m$ for som constant c. The m are deductively closed with respect to $\vdash_s$, for if $A \notin m$ and $m \vdash_s A$ then $\sim A \in m$ and so $m \vdash_s \sim A$, but we have assumed that m is consistent.

The deduction[13]

$$\dfrac{\dfrac{\sim\forall x A \qquad \dfrac{\dfrac{\dfrac{\cancel{x = x}^2 \quad \dfrac{\dfrac{\cancel{A}^1}{\bot}\ \textit{restart}}{\sim A}\ \supset I(1)}{\exists x \sim A}\ \exists I}{A}\ \textit{restart}}{\forall x A}\ \forall I(2)}{\bot}\ \supset E}{\exists x \sim A}\ \bot E$$

shows that for any such m, if $\sim\forall x A \in m$ then $\sim A[x/c] \in m$ for some c. It is not complicated to show that every set consistent with respect to $\vdash_s$ can be extended to a set m. For example, the consistency of the rule $\exists E$ means that it is consistent, indeed conservative, to extend every consistent Γ to Γ' so that for every $\exists x A \in \Gamma'$ there is c such that $A[x/c], c = c \in \Gamma'$. Therefore if A is in all m so extended that contains Γ, then A is a deductive consequence of Γ.

[13] A similar deduction can be constructed using PIP and not the restart rule.

Each m may be considered a world, the elements of which are nonempty sets of constants $\{a, b : a = a \in m \ \& \ a = b\}$. That is, if a is a constant and $a = a \in m$ then a^I is the set of constants $\{b : a = b \in m\}$. We may take the set of all such equivalence classes of constants as the domain D of the entire model.

The extension if each atomic predicate F^I is the set of all $\langle a_1^I \dots a_n^I, m\rangle$ where the a_i are constants and $Fa_1 \dots a_n \in m$.

A valuation is taken as usual, as an assignment of variables and constants to elements of D such that $a^I = a^v$ if a is a constant.

If $\langle t_1^v \dots t_n^v, w\rangle \in F^I$ then, by the definition of F^I, $Fa_1 \dots a_n \in m$ for some constants $a_i \in t_1^v$.

If $Ft_1 \dots t_n \in m$ then $t_i = t_i \in m$ for each term t_i by $=I$. By $\exists I$ and then and the definition of each m, $Fc_1 \dots c_n \in m$ for some constants $c_1 \dots c_n$. Therefore we any valuation that sets each t_i^v to c^I is such that such that $\langle t_1^v \dots t_n^v, w\rangle \in F^I$.

The accessibility relation on the m is defined as in 10.2.2

$$mRm' \text{ iff } \{B : A \to B \in mA \in m'\} \subseteq m'$$

we can verify, as we did in 10.2.2 that R is reflexive and transitive.

The actual world m_a by taking some world m such that $x = x \in m$ for every variable x. The canonical valuation v_c sets $x^v = a^I$ for every variable x s.t. $x = a \in m_a$

It is now fairly easy to show that we have constructed a model M for $\vdash$. We may now set that $M \vDash_m^v A$ iff $A' \in m$ where A' results by replacing every variable x in A by some constant $c \in x^v$. We must verify that this yields the correct truth conditions.

If A is atomic this has already been shown above.

$M \vDash_m^v {\sim}A$ iff ${\sim}A' \in m$, iff $A' \notin m$ (by definition of m) iff, by induction hypothesis, $M \nvDash_m^v A'$.

$M \vDash_m^v A \wedge B$ iff $(A \wedge B)' \in m$ iff, by $\wedge I$ and $\wedge E$, $A' \in m$ and $B' \in m$ iff, by induction hypothesis, $M \vDash_m^v A$ and $M \vDash_m^v B$.

$M \vDash_m^v \exists x A$ iff $\exists x A' \in m$ iff $A'[x/c] \in m$ for some c s.t. $c = c \in m$ by the definition of m and the $\exists I$ rule iff, by ind. hyp. $M \vDash_m^{v'} A'$ on some x-alternate valuation v' to v.

We can show that

$$A \to B \in m \text{ iff } B \in m' \text{ for every } m' \text{ s.t. } mRm'.$$

by reasoning similar to that of page 138.[14]

[14]It follows in the left-right direction by the definition of R. Now let $M' = \{m' :$

Thereby we can show that $M \vDash^v_m A \to B$ iff $M \vDash^v_{m'} B$ for every m' s.t. mRm' and $M \vDash^v_{m'} A$.

The other cases can be shown similarly, or by proving equivalences to formulae expressible using only $\exists, \sim, \to, \wedge$.

Finally, our choice of m_a is such that $x = x \in m_a$ for every variable x. Consequently, for every variable x, there is a constant c such that $x = c \in m_a$. Thus v_c assigns an element of m_a to every variable of the language.

Now, if $\Gamma \nvdash A$ then $\Gamma' = \Gamma \cup \{\sim A\}$ is consistent.[15] Since $\vdash x = x$ we can extend Γ into a maximal consistent set $m_{\Gamma'}$ with respect to $\vdash$ where $x = x \in m_{\Gamma'}$ for all variable x, and, since t_s is weaker we can find $m_{\Gamma'}$ in the model described above. We can then set the actual world $m_a = m_{\Gamma'}$ above. Since $\sim A \in m_a$ we have that $M \vDash^v_{m_a} \sim A$, and we have found a model where Γ is true and A is not. So if A is true in all models where Γ is true then $\Gamma \vdash A$. This concludes the (somewhat sketchy) completeness proof.

11.3.6 A final comment

In the system I have given, every $a = a$ is false if a does not refer (and all variable refer). This might not be entirely appropriate. We can modify the logic easily so that $Ft_1 \dots t_n$ is false if t_i does not refer, but $t_i = t_i$ is necessarily true. For this we need a dummy unary atomic predicate E. The rules for equality revert to the normal classical rules, in particular $=I$ is

$$\frac{}{t = t}$$

$A \in m'$ & $mRm'\}$. Suppose further that $B \in m'$ for every $m' \in M'$. It follows that $\{C : A \to C \in m\} \cup \{A\} \vdash_s B$. But then $\{A \to C : A \to C \in m\} \cup \{A \to A\} \vdash A \to B$.

$$\dfrac{A \to C_1 \dots A \to C_n \qquad \begin{array}{c} \cancel{A}^1 \quad \begin{array}{c} \cancel{(A \to C_1)}^1, \dots, \cancel{(A \to C_n)}^1, \cancel{A}^1 \\ \vdots \ {\to}E \\ C_1 \dots C_n \end{array} \\ \vdots \\ B \end{array}}{A \to B} \ {\to}I(1)$$

is the deduction, notice that the side conditions on $\to I$ entail, effectively, that $\{A \to C_1 \dots A \to C_n\} \vdash_s C$.

$A \to A$ is a theorem of the logic and $\{A \to C : A \to C \in m\} \subseteq m$, therefore $A \to B \in m$ by the deductive closure of m. This shows the right-left direction.

[15] As if $\Gamma, \sim A \vdash \bot$ then $\Gamma \vdash A$ (using restart or PIP).

and E has an introduction rule:

$$\frac{B}{E(t)}\ EI$$ where B is an atomic formula in which t is free. If t is a variable then B need not be present.

In effect, E is an existence predicate, but it need not be interpreted as predicating of nonexistent objects since if $\sim E(t)$ then any atomic predication containing t is false. E is more like a reference predicate, it does not entail the presence of nonexistent objects.

11.4 Conclusion

We can produce a harmonious system of rules that allows for non-referring terms, when used in conjunction with the rules for strict implication we obtain a very subtle modal logic a semantics for which involves possible worlds with variable domains. This seems all very much on the right track. Furthermore the distinction between weak and strong rules allows us to achieve another subtlety, we can have $x = x$ a theorem of the logic without $\Box(x = x)$ being a theorem. Thus we can account for the analyticity of a sentence like 'I am me' or 'I exist' without that entailing its necessity (by finding more rules for terms like 'here' and 'now' I expect that we can account for sentences such as 'I am here now').

Chapter 12

$\perp$ and the ex falso rule

Perhaps the least intuitive of the rules suggested so far is the elimination rule for $\perp$. Indeed, the ex falso rule is so unintuitive that it may be taken as evidence that any logic containing it is not the logic of our reasoning (i.e. not the logic we define). This would be evidence against my thesis as part of my argument (that classical logic is analytic) is that we actually use classical logic in our reasoning. I discuss the ex falso rule here and try to make it seem less unintuitive. But my main argument for accepting it is this:

- The logic we use has a rich and simple truth conditional semantics. Alternatively put, the properties and relations that hold between propositions are not obscure and are easy for us to analyse. In particular we can provide simple and intuitive accounts of what propositions are that are at least of instrumental value (e.g. possible world semantics, no matter how implausible it may seem, is at least of instrumental value in the practice of philosophy).

- However it seems that the best alternative logics that do not contain the ex falso rule have very obscure, trivial, or unhelpful semantics. I refer here to the search for a simple semantics for relevance logics, or a simple theory of propositions for relevance logic.[1]

- Therefore the logic we use is such that everything follows from a contradiction (i.e. it does contain the ex falso rule).

[1] Semantics for relevance logic are known, but they are by no means intuitive and require significant logical ability and experience to understand. I submit that they are of little value in furthering our understanding of the nature of propositions and truth conditions.

12.1 $\perp$

Traditionally $\perp$ is regarded as denoting an absurdity, or something that is disbelieved by definition. Treating $\perp$ as shorthand for a contradictory pair of propositions $\{A, \text{not-}A\}$ is not enough for the purposes of this thesis unless we add that contradictions are absurd or disbelieved by definition (so that the absurdity of contradictions is analytic). We might argue that the absurdity of contradictory pairs is part of the definition of negation, this is not unacceptable although a little inelegant. I think an equivalent and more natural strategy is to regard $\perp$ as a formula that is absurd or disbelieved by definition, and show how the rules for negation entail that $\{A, \text{not-}A\} \vdash \perp$. In light of this I shall try to regard $\perp$, not as denoting any contradiction, but as an atomic formulae which is necessarily absurd or disbelieved.

The problem then is whether such a formulae should support the famous *ex falso quod libet* elimination rule:

$$\frac{\perp}{B}$$

which allows the derivation of anything from absurdity and consequently from a contradiction. It is hard, if not impossible, to find uses of this rule explicitly, which has lead many to believe that it is not a correct rule of inference and should not be part of our primitive logic.

12.2 Arguments for ex falso

12.2.1 Lewis' deduction

Here is C.I. Lewis' famous argument that anything follows from a contradiction:

$$\dfrac{\text{not-}A \quad \dfrac{A}{A \vee B}\,\vee I}{B}\,DJ$$

The final step DJ is *disjunctive syllogism*, an argument pattern which is undeniably commonplace in common reasoning. Perhaps less common is the rule $\vee I$ which is what seems suspect initially to many people. However it is hard to fault as $A \vee B$ is indeed true if A is.[2] We can justify ex falso using this argument if we regard $\perp$ as a shorthand for a contradictory pair (any contradictory pair). We might then even use this result to explain why

[2]It is enough to verify that A in order to verify that $A \vee B$.

contradictions are analytically absurd (and disbelieved on principle), they entail everything.

Equally famous is relevance logic which is motivated to rebut Lewis' argument. The objection is that the argument is circular, the circularity becomes apparent when we try to justify disjunctive syllogism. The elimination rule for disjunction is:

$$\frac{A \vee B \quad \begin{array}{c}\cancel{A}\\ \vdots \\ C\end{array} \quad \begin{array}{c}\cancel{B}\\ \vdots \\ C\end{array}}{C}$$

it is used to justify disjunctive syllogism

$$\frac{A \vee B \qquad \dfrac{\dfrac{\cancel{A} \quad \text{not-}A}{\bot}}{B} \qquad \cancel{B}}{B}$$

and the application of the disputed ex falso rule has been revealed. In order to justify the ex falso rule Lewis' argument used disjunctive syllogism which itself assumed the ex falso rule. We might try to repair this by arguing that disjunctive syllogism is an elimination rule for disjunction and is not derived. This is a plausible suggestion, disjunctive syllogism seems a basic rule of inference, perhaps it is basic enough to be definitional of $\vee$. We might add two extra elimination rules for disjunction:

$$\frac{A \vee B \quad \begin{array}{c}\cancel{A}\\ \vdots \\ \bot\end{array} \quad \begin{array}{c}\cancel{B}\\ \vdots \\ C\end{array}}{C} \qquad \frac{A \vee B \quad \begin{array}{c}\cancel{A}\\ \vdots \\ C\end{array} \quad \begin{array}{c}\cancel{B}\\ \vdots \\ \bot\end{array}}{C}$$

and now disjunctive syllogism can be validated without use of ex falso. Such a response is of little help to this thesis as those elimination rules cost us normalisation (harmony) unless we have also the ex falso rule. Unless we have ex falso as a primitive (underived) inference rule, a deduction of B from a contradiction must involve the introduction and subsequent elimination of a connective (disjunction) which cannot be reduced to a normal form. Perhaps this is more the worse for the demand for normal forms, but the roundabout nature of Lewis' argument makes it seem like a sophistry of some sort.

Actually, the argument need not utilise disjunction at all and may avoid the charge of circularity by reformulating the argument in terms of conjunction and negation. For example the following deduction

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\text{not-}A \qquad \dfrac{\cancel{A\wedge\text{not-}B}}{A}}{\bot}}{\text{not-}(A\wedge\text{not-}B)} \qquad \dfrac{A \quad \cancel{\text{not-}B}}{A\wedge\text{not-}B}}{\bot}}{\text{not-not-}B}}{B}$$

does not assume the ex falso rule. An application of double negation elimination (or some other rule that entails it) is required. Denying double negation elimination yields little relief as a similar argument will deduce the negation of every formula from a contradiction which seems to me to be as problematic as the ex falso rule. The use of conjunction in this argument turns out to be extremely problematic for relevance logic and, it seems to me, that relevance logics have not yet achieved an adequate treatment of conjunction.

Whatever is to be made of the Lewis argument it is not enough for me to use it as a justification of the ex falso rule. I am investigating how much logic we can get when operating under the normalisation constraint, and the Lewis argument is not in normal form, it can be reduced to normal form only if the ex falso rule is taken as a primitive deduction rule. I must therefore seek an alternative justification of $\bot$.

12.2.2 Dummett's justification

Dummett provides two interpretations of $\bot$,

1. $\bot$ is the set of all atomic formulae ([**?**])
2. $\bot$ is some false atomic formula typically $0 = 1$ ([**?**]).

The first interpretation of $\bot$ legitimates the following restricted ex falso rule:

$$\dfrac{\bot}{B}\ \bot E' \qquad \text{Provided } B \text{ is atomic.}$$

an easy induction on the degree of B shows how $\bot$ deduces any B by a sequence of introduction rules.[3]

[3] For example, if B is atomic then $\bot \vdash B$ immediately by $\bot E'$; if $B = C \wedge D$ then $\bot \vdash C$ and $\bot \vdash D$ by induction hypothesis and so $\bot \vdash B$; if $B = \neg C$ then since $\{\bot, C\} \vdash \bot$,

The second interpretation seems circular at first sight. However, in the case of (certain views concerning) arithmetic it differs very little, if at all, from the first. To see this note that atomic formulae in the language of arithmetic are all of the form $t = t'$. With the axioms of Peano arithmetic without the axiom that $s(x) \neq x$ (s is the successor function) we can use $0 = 1$ to derive every atomic formula (open or closed) from $0 = 1$. Thus, at least in the case of the language of arithmetic, there is a close relation between the two interpretations of $\perp$. Since we do not speak only in arithmetic this observation does not apply and the second interpretation does not justify the ex falso rule, we may take it for granted that $0 = 1$ is false necessarily, but we do not, in general think that $0 = 1$ entails that the sun is shining.[4]

The first interpretation, that $\perp$ is all atomic formulae, is unproblematic from the point of view of justifying ex falso. The problem arises when we ask

---

it follows by $\neg I$ that $\perp \vdash \neg C$; if $B = C \vee D$ then $\perp \vdash C$ thence $\perp \vdash B$ by $\vee I$; finally if $B = C \rightarrow D$ then $\perp \vdash D$ and so $\{\perp, C\} \vdash D$ thence $\perp \vdash A \rightarrow B$. The argument is unchanged if we uses classical negation and implication symbols instead.

[4]Although perhaps this is because we have not thought enough, consider this proof that p.

> either p or not-p, if p then there is nothing left to prove. On the other hand, if not-p then the number of things such that p is 0, but since $0 = 1$ the number of things such that p is 1, so there is something s.t. p, therefore p.

an argument not relying on the law of excluded middle is perhaps constructible using extra rules relating to 'the number of things such that p'. I am not sure what to make of this argument. If it cannot be faulted, and I see no fault, then it shows that ex falso is true provided that $\perp$ is short for $0 = 1$ and appears in a fairly strong arithmetic system. However it seems wrong to hold that every sentence of the form not-A (e.g. the sun is not shining) contains an arithmetical sense to it (e.g. if the sun is shining then $0 = 1$). In general we do not introduce negation on the basis of a deduction that $0 = 1$.

Diversions through maths seem to me to be the quickest of the intuitive derivations of anything from a contradiction. Here is another example:

> Suppose that p and not-p. Now there is exactly 1 thing identical to me and so, virtue of p, there is (exactly) one thing identical to me such that p. But since not-p then there are 0 things such that p and so 0 things that are identical to me such that p. So the number of things that are identical to me and such that p is both 1 and 0 and so $1 = 0$. But, as argued above, anything follows from '$1 = 0$'.

So, if $\perp$ is represents any contradiction, and the above arguments are acceptable, and we have a priori enough maths to be claims like 'the number of things such that...' intelligible, then the ex falso rule is acceptable (and we need not regard $\perp$ as short for $0 = 1$). However I find the detour through mathematical claims worrying, it is against the spirit of normalisation (which I use as the basis for analyticity). We should expect there to be more direct arguments or intuitions that anything follows from a contradiction if ex falso is indeed an analytic law of logic.

if, in actual practice, we really believe negations on the basis of a deduction of every atomic formula. When I judge that there is no glass on the table in front of me I do so by failing to find one, this is not implausibly modelled by an application of the introduction rule for negation[5]

$$\dfrac{\begin{array}{c}\cancel{A}\\ \vdots\\ \perp\end{array}}{\text{not-}A}$$

but not if $\perp$ is all atomic formulae, certainly I do not try to verify that the sun is shining in order to show that there is no glass on the table.

Furthermore if $\perp$ is all atomic formulae then its meaning differs when the language is extended, a new $\perp$ must be introduced when new atomic formulae are added to the language (e.g. if a new word is introduced meaning 'snow' then we have a new atomic formulae which means that snow is white).

Dummett's interpretations of $\perp$ are quite natural in the context of a language for mathematics but not for natural language.

12.2.3 An definitional justification

It may be thought that we can justify the ex falso rule in exactly the same way we intend to justify the other logical connectives: we specify that the ex falso rule (or rules that entail it) is part of the *definition* of $\perp$.

Usually, the elimination rule for $\perp$ is the ex falso rule and $\perp$ has no introduction rule. We now need only check that the introduction and elimination rules are in harmony. But since $\perp$ has no introduction rule, it can never be introduced so we need never worry about it being introduced and then eliminated. Certainly in all the logics I consider in this thesis, the ex falso rule causes no problems for proving normal form theorems.

This may show that there are no logical grounds for rejecting the ex falso rule, but it does not make ex falso any more plausible as a rule to adopt, and does not help to allay any fear that it is a rule that we do not actually adopt. If we are to say that ex falso is a definitional rule, we must give some argument that it is used as a definition. But this is precisely what is being doubted, it seems that the ex falso rule is never used (moreover, it seems irrational to use it).

It might be argued that we should always adopt the strongest possible elimination rule that is in harmony with the introduction rule. But this

[5] Where the assumption of A is implicit in the attempt to find the glass and is discharged when I terminate the 'search'.

requirement is too strong, for example many atomic formulae other than $\perp$ also have no introduction rules, but we would not wish to have a rule

$$\frac{A}{B}$$

for every such atomic formula. No doubt a subtle distinction could answer this little problem, but the more general problem is that sometimes we may not wish the strongest possible elimination rule.[6]

Furthermore what rule is the strongest may depend on what other connectives have been defined and even the way they are formulated. Perhaps a condition of the form 'adopt the strongest possible elimination rule' can be made more clear and more general, but I shall not develop it here.

I believe it is not enough to show that the rules for a logical connective are in harmony, for this shows only that a definition can be made. It does not show that the definition has been made or must be made, and in the case of the rules for $\perp$ this is precisely what is in contention.

I think we can resolve the difficulty here by distinguishing between two sorts of rules: direct inference rules and admissible inference rules. An admissible rule is a meta-level rule that says 'if there is a deduction of the premises then there is also a deduction of the conclusion'. For example, here is the ex falso rule understood as a direct inference rule:

[6] For example consider these rules, which we may add instead of the rules for strict implication (matters are slightly more complex if we have rules for strict implication as well):

$$\frac{A}{\Diamond A}\ \Diamond I$$

$$\frac{\Diamond A \quad B_1 \ldots B_n \qquad \begin{array}{c}\cancel{A}, \cancel{B_1} \ldots \cancel{B_n}\\ \vdots \\ B\end{array}}{B}\ \Diamond E1$$

Provided (1) B is *backwards-strong* (2) each B_i is strong (3) B depends only on $A, B_1 \ldots B_n$

and where, for any A, B, $\Diamond A$ is backwards-strong and $\perp$ is backwards strong. Furthermore, we must add that any incomplete applications of restart the premise of which is backwards-strong is not a weak rule.

With these additions we can show that $\Diamond A$ has the familiar Kripke semantics of $S4$ where $\Diamond A$ is true iff A is in some accessible possible world. I shall not prove this here.

We should not rule, out of hand, that $\Diamond$ is not a legitimate or useful connective to define. But clearly the elimination rule for $\Diamond$ are not as strong as it could be:

$$\frac{\Diamond A}{A}$$

is the strongest rule in harmony with the introduction rules.

> From $\perp$ (you may) infer A

and as an admissible inference rule:

> If you have an inference of $\perp$ then you *can* obtain also an inference of A.

More formally the direct inference rule for $\perp$ is this:

$$\frac{\perp}{A} \perp E$$

Whereas the admissible rule is this

$$\frac{\Gamma \vdash B}{\Gamma \vdash A}$$

Now, it is not hard to show, that in a logic with no rules specifically for $\perp$ then the ex falso rule is indeed an admissible rule (even though there is no direct inference rule corresponding to ex falso).

If our general theory of definition-by-inference-rules requires the inference rules to be direct inference rules then we have problems arguing that a logic with the ex falso rule is the logic we actually use. But if our general theory of definition-by-inference-rules allows also that admissible rules can count as definitional, then perhaps we will have an easier time arguing that classical logic (with its ex falso rule) is a logic we actually define.[7] On this understanding of the ex falso rule, it is not a rule that says that we may infer any A from a contradiction ($\perp$). Instead it is a general condition on our logic that any deduction of a contradiction may as well be a deduction of anything. Put more loosely, the ex falso rule (interpreted as a definitional admissible rule) is a promise never to accept a contradiction.

If we interpret the ex falso rule as an admissible rule then we may accept that it is irrational to infer A directly from $\perp$ (as there is no inference rule that allows it). However, we can still regard the ex falso rule as a rule that

[7] In Gentzen's original writings on the idea that inference rules can define logical constants, the inference rules he used are formulated in his sequent calculus. A sequence calculus may be thought of as a deduction system for the logical consequence relation derived from some other deduction system. For example one of the sequent rules for conjunction is this:

$$\frac{\Gamma \vdash A \quad \Gamma \vdash B}{\Gamma \vdash A \wedge B}$$

which does not say that we may infer $A \wedge B$ from both A and B, it says something more general (that if there are two deductions of A and of B, then *there is also* a deduction of $A \wedge B$, nothing is said about how this new deduction is obtained).

defines our basic logic, and as a crucial rule in making our logic complete for the familiar truth conditional semantics.

I think it is plausible that we do treat the ex falso rule as admissible. I use as evidence remarks such as

> If you believe that then you'll believe anything!
>
> If we accept that, then we might as well accept this: [insert ridiculous proposition here]!

This sort remark may be more than just a rhetorical device.

So in conclusion, we may use our theory of definition-by-inference-rule to account for the meaning of $\bot$ (and justify the ex falso rule), if we allow that admissible rules may be used to define a logical connective. I think it is not entirely implausible that such a story can be made to work at least for reasoning from contradictions (or $\bot$). I tentatively conclude therefore, that we may resolve the difficulties with the ex falso rule (as part of a theory of how we actually reason) by regarding it as an admissible rule rather than a direct inference rule.

To be fair to those who are not convinced by the justification of ex falso I have just proposed, I now turn to a discussion of what a system of inference rules without the ex falso rule might look like.

12.3 Life without ex falso

I shall reject out of hand paraconsistent logics which allow contradictions to be true, or at least do not have the law of non-contradiction as a theorem. As far as I am concerned, regardless of the status of the ex falso rule I have already shown that the meanings of the terms entail the truth of $\ulcorner$not-$(A \wedge$ not-$A)\urcorner$. Here is the deduction (valid in intuitionistic logic):

$$\dfrac{\dfrac{\dfrac{\cancel{A \wedge \neg A}}{A}\wedge E \quad \dfrac{\cancel{A \wedge \neg A}}{\neg A}\wedge E}{\bot}\bot E}{\neg(A \wedge \neg A)}\neg I$$

12.3.1 Relevance logic: implication

Relevance logic is then worthy of some extra discussion. Every line in a relevance deduction is annotated by a set of labels (integers will do as labels), these labels are transmitted down the branches of the deduction and indicate what assumptions were required for the deduction of each formula, for example:

$$\frac{1 : A \quad 2 : A \to B}{1, 2 : B}$$

is an application of Modus Ponens. The two premises have the sets labels $\{1\}$ and $\{2\}$ and the conclusion has the union of the labels of its premises (here, the set of labels $\{1, 2\}$).

The rules for implication in relevance logic are:

$$\frac{\alpha : A \quad \beta : A \to B}{\alpha \cup \beta : B} \qquad \frac{\begin{array}{c} n : \cancel{A} \\ \vdots \\ \alpha \cup \{n\} : B \end{array}}{\alpha : A \to B}$$

notice that in the introduction rule we can discharge A only if the label we initially assign A is one of the labels of B, the idea being that $A \to B$ is deducible only if A was used in the deduction of B.

For example we cannot deduce $A \to B$ from B by empty discharging A, for, as the name suggests, A is empty discharged when it was not used in deducing B. Thus $A \to (B \to A)$ is not a theorem of relevance logic. The conclusion of any deduction in relevance logic with no labels attached to it is a theorem. For example:

$$\frac{\frac{\frac{\frac{\frac{1 : \cancel{A} \quad 2 : \cancel{A \to B}}{1, 2 : B} \quad \frac{1 : \cancel{A} \quad 3 : \cancel{A \to (B \to C)}}{1, 3 : B \to C}}{1, 2, 3 : C}}{2, 3 : A \to C}}{3 : (A \to B) \to (A \to C)}}{[A \to (B \to C)] \to [(A \to B) \to (A \to C)]}$$

shows that $[A \to (B \to C)] \to [(A \to B) \to (A \to C)]$ is a theorem of relevance logic. Also, in relevance logic $A \to (B \to C) \vdash (A \to B) \to (A \to C)$ as there is a deduction of $\alpha : (A \to B) \to (A \to C)$ from $\alpha : A \to (B \to C)$, that is, any assumptions that deduce $A \to (B \to C)$ also deduce $(A \to B) \to (A \to C)$

12.3.2 Problems with relevance logic: conjunction and disjunction

In order to block the two Lewis arguments, in particular the one utilising conjunction rather than disjunction, relevance logic places the following restrictions on conjunction and disjunction (Greek letters are for sets of labels.[8] For brevity let α, β be short for $\alpha \cup \beta$ and α, n be short for $\alpha \cup \{n\}$:

$$\frac{\alpha : A \quad \alpha : B}{\alpha : A \wedge B}\wedge I \qquad \frac{\alpha : A \vee B \quad \begin{array}{c}\alpha : \cancel{A}\\ \vdots \\ \alpha, \beta : C\end{array} \quad \begin{array}{c}\alpha : \cancel{B}\\ \vdots \\ \alpha, \beta : C\end{array}}{\alpha, \beta : C}\vee E$$

In other words, we can introduce $A \wedge B$ only if both conjuncts are derived from the same assumptions, and we may eliminate $A \vee B$ only if we can derive C from A and from B using the same additional assumptions. These restrictions are to block the addition of irrelevant assumptions. The introduction rule for negation is this:

$$\frac{\begin{array}{c}n : \cancel{A}\\ \vdots \\ \alpha, n : \bot\end{array}}{\alpha : \neg A}$$

where n is a fresh integer that labels no other assumption.

Now consider this deduction:

$$\frac{1 : A \vee B \quad \begin{array}{c}1 : \cancel{A} \quad 2 : \neg A\\ \vdots \\ 1, 2 : \bot\end{array} \quad \begin{array}{c}1 : \cancel{B} \quad 3 : \neg B\\ \vdots \\ 1, 3 : \bot\end{array}}{1, 2, 3 : \bot}$$

in which we allow the elimination of $A \vee B$ even though the two minor premises do not depend on the same assumptions. But now both $\neg A$ and $\neg B$ have become relevant to the deduction of $\bot$ (as its labels indicate) we may then finish the deduction by introducing a negation:

$$\frac{\dfrac{1 : A \vee B \quad \begin{array}{c}1 : \cancel{A} \quad 2 : \neg A\\ \vdots \\ 1, 2 : \bot\end{array} \quad \begin{array}{c}1 : \cancel{B} \quad 3 : \cancel{\neg B}\\ \vdots \\ 1, 3 : \bot\end{array}}{1, 2, 3 : \bot}}{1, 2 : \neg\neg B}$$

and a further application of double negation elimination (a usual part of relevance logic) yields disjunctive syllogism and from there the ex falso rule.

[8]There are different systems of relevance logic, I present here one of the stronger (and more sensible) elimination rules for disjunction.

A similar problem arises with conjunction, suppose we allow that the premises of conjunction introduction have different labels, then we can use conjunction to add irrelevant premises.

$$\dfrac{\dfrac{1 : A \quad 2 : B}{1, 2 : A \wedge B}}{1, 2 : A}$$

In this deduction conjunction is introduced and then eliminated in order to add an extra label to A: the label for B. But B might not be relevant to A.

The restrictions on conjunction introduction and disjunction elimination block such deductions, but the price is high. Firstly such restrictions seem highly unnatural: we should be able to introduce a conjunction whenever we have derived each conjunct. Secondly $\{\neg(A \wedge B), A, B\}$ and $\{A \vee B, \neg A, \neg B\}$ are not contradictory sets with these restrictions in place, but they ought to be.[9] These problems have caused some to reject relevance logic as of merely technical interest (e.g. Hanson [**?**]). But those are not the only possible rules, I shall now present a deduction system for relevance logic that does not suffer from these problems.

12.3.3 Tennant's relevance logic

Before I present my solution to these problems I wish to discuss another solution offered by Tennant in [**?**] and [**?**]. Tennant observes the high cost of such restrictive rules for conjunction and disjunction (see the labelled rules above), and also rejects the famous relevance logic denial of disjunctive syllogism! Tennant rejects the Lewis 'proof' of the ex falso rule by rejecting the transitivity of deduction. So, in Tennant's system, we may have a deduction from A to B and another from B to C, but then we may not be able to obtain from these a deduction from A to C.

Tennant's logic is effectively this

- It contains all of the familiar rules for intuitionistic logic except the absurdity elimination rule, and has additionally the rule PIP (to make it classical).
- It has an extra condition banning any empty discharging (this affects the rules PIP, $\rightarrow I$, $\vee E$ and $\exists E$.
- It has a global condition that all deductions must be in normal form

[9] Note also that $\{A \wedge B, A \rightarrow C, B \rightarrow D\} \nvdash C \wedge D\}$, and $\{A \vee B, A \rightarrow C, B \rightarrow D\} \nvdash C \vee D\}$ in relevance logic, which is quite a failing.

So for example although these are valid deductions:

$$\frac{A \quad B}{A \wedge B}\wedge I \qquad \frac{A \wedge B}{A}\wedge E$$

this is *not* a valid deduction:

$$(\dagger) \quad \cfrac{\cfrac{A \quad B}{A \wedge B}\wedge I}{A}\wedge E$$

I have some issues with Tennant's relevance logic, I give two here:

- Deduction *is* transitive. There is nothing wrong with (†) as a piece of reasoning except that it need not be so long. I argue for this claim only by appeal to our practices in reasoning. If we have reason to believe that A deduces B and that B deduces C we rarely (if ever) check that we can reason from A to C in normal form before we conclude that A deduces C. Mathematical practice would be almost impossible if we had to redo all previous lemmas and theorems just so we can derive a corollary. We frequently reason out of normal form (especially when chains of reasoning are done partly by one reasoner and partly by another).

- In Tennant's logic the object language conditional does not match entailment in the logic. We can deduce $A \rightarrow C$, in normal form, from $A \rightarrow B$ and $B \rightarrow C$. We can even deduce $(A \rightarrow B) \rightarrow [(B \rightarrow C) \rightarrow (A \rightarrow C)]$. So in a sense the conditional is transitive (or at least the conditional 'thinks' it is transitive).[10] But then $A \rightarrow B$ is not the object level analogue of the meta-level 'there is a deduction/entailment from A to B'. This makes it hard to use Tennant's logic together with my account of implicit definition in chapter 14, for my account requires the object level connectives to be analogues of meta-level phrases and constructions.

To conclude then, Tennant's relevance logic, as Tennant shows in [**?**] possesses substantial logical power despite its failure of transitivity. Nevertheless, I object that it will not do as the basis of a theory of analyticity. The primary objection here is that it does not match actual reasoning (reasoning is transitive). In section 12.4.5 I suggest that any relevance logic will suffer from this problem (of not being a good account of human reasoning) for semantic reasons.

[10] Actually it is not, for we may have deductions of $A \rightarrow B$ and $B \rightarrow C$ but not be able to convert them into a deduction of $A \rightarrow C$ without breaching the restriction the deductions only be in normal form.

12.3.4 My diagnosis of the problem

It would be an injustice to relevance logic not to attempt further to overcome the difficulty of formulating relevance logic so it is plausible that we actually use it. The idea behind the relevant conditional is intuitive: $A \to B$ is true when A is relevant to the conclusion B.

As seen above we may use conjunction introduction (and disjunction elimination) to add irrelevant premises to a deduction. We deduce A from assumptions [labelled by] α, and B from assumptions β and thence $A \wedge B$ from the union of these assumptions:

$$\frac{\alpha : A \quad \beta : B}{\alpha, \beta : A \wedge B}$$

the problem is that we might later use $A \wedge B$ to conclude A and mistakenly be led to believe that the assumptions β are relevant to the deduction of A. For example:

$$\cfrac{\cfrac{\cfrac{\cfrac{1 : \cancel{A} \quad 2 : \cancel{B}}{1, 2 : A \wedge B}}{1, 2 : A}}{1 : B \to A}}{A \to (B \to A)}$$

is a deduction of a formula relevance logicians do not wish as a theorem.[11] We must find some way of blocking the introduction of $B \to A$ in the deduction above. Notice this deduction with a single premise

$$\begin{array}{c} A \wedge B \\ \vdots \\ C \end{array}$$

may be viewed only as a deduction of C from a premise $A \wedge B$. But a similar deduction of C from more than one premise

$$\begin{array}{c} A, B \\ \vdots \\ C \end{array}$$

is in a sense at least three deductions. It is a deduction of C from A in the context of B, a deduction of C from B in the context of A, and a deduction of C from A and B. But it might not be relevant that A and B are separate premises, for example if they are immediately used to introduce $A \wedge B$.

[11]If $\neg B$ is short for $B \to \perp$ then the theorem entails that $\perp \to \neg B$ for any B, which together with double negation elimination yields that $\perp \to B$ for any B.

This suggests the following addition to the conjunction introduction rule:

$$\frac{A \quad B}{A \wedge B}$$ Provided the assumptions on which A and B depend are discharged together or not at all.

and the deduction above of $A \to (B \to A)$ is blocked as $A \wedge B$ is introduced from assumptions A and B, which are subsequently discharged at different points in the deduction.

We need to add a mechanism of labelling assumptions so we may give a clearer account of simultaneous discharge. We already use a minor labelling system in our practice of adding superscripts to formulae at top nodes of deductions, it just so happens that this is enough.

12.4 A less problematic relevance logic

I shall present a natural deduction system here for a relevance logic, it is a simplification of a more subtle natural deduction system which I shall not describe in this thesis. I do not know exactly which axiom system deduces exactly the theorems of the natural deduction system I am about to present (I suspect it is $\Box R$ without the axiom for distribution of conjunction across disjunction).[12] I present the relevance logic here just as an example of how we can obtain harmonious natural deduction systems for relevance logics without such heavy use of labels and with less restrictive rules for conjunction and disjunction.

- Say that occurrences of $A_1 \dots A_n$ in a Prawitz tree *depend on the same applications of restart* when for every incomplete application of restart on which A_i depends, there is an incomplete application on which A_j depends with the same premise (for any $i, j \leq n$).

12.4.1 The rules

Restart

Firstly the restart rule must be restricted:

$$\frac{A}{\perp}\ \mathit{restart}$$ Provided that A is occurs in the Prawitz tree below $\perp$

we may also call this restricted (or, relevant) restart $\perp I$.

[12] See J. Michael Dunn, *Relevance Logic and Entailment*, in [?] Volume III, chapter 3.

Connectives

The side conditions on the following rules are somewhat long, although the concept behind them is fairly simple. It is possible to shorten the side conditions by means of a more compact language, but there is no point here. Also the side conditions will almost completely disappear if I give the following rules in a multiple conclusion logic (indeed, the normalisation proof below will be considerably simpler in a multiple conclusion logic). I present the logic here as a single conclusion calculus to keep it in the spirit of the other logical systems of this thesis.

- In the following inference rules a premise such as this:

$$\begin{array}{c} [\cancel{A_1}^m] \ldots [\cancel{A_n}]^m \\ \vdots \\ B \end{array}$$

  represents any Prawitz tree some of the top nodes of which are crossed out occurrences of A with a superscript m.[13]

Implication is enhanced so that it can take multiple antecedents. That is ( )$\rightarrow$, or $\rightarrow$ for short, is a connective that has an arbitrary finite string of antecedents and one consequent. So if $A_1 \ldots A_n, C$ are formulae then so is $(A_1 \ldots A_n) \rightarrow C$.[14]

Here is the introduction rule for $\rightarrow$ (each n_i is an integer):

$$\dfrac{\begin{array}{c} [\cancel{A_1}^m] \ldots [\cancel{A_k}^m] \\ \vdots \\ B \end{array}}{(C_1 \ldots C_l) \rightarrow B} \rightarrow I(m) \quad \text{Where the } A_1 \ldots A_k \text{ are among } C_1 \ldots C_l.$$

This rule is formulated so that every rule application must be have the additional label m, and each A_i must appear at top nodes of the Prawitz tree (with a superscript m). Since the rule application must have the additional label m, each A_i is discharged by that rule application (as no two rule

[13]This:

$$\begin{array}{c} \cancel{A_1} \ldots \cancel{A_n} \\ \vdots \\ B \end{array}$$

represents any Prawitz tree which may or may not have some crossed out occurrences of A at top nodes.

[14]Interpret $(A_1 \ldots A_k) \rightarrow B$ as ⌜If A_1 *and if* A_2 *and if*... A_k then B⌝.

applications can have the same additional label). Here is the elimination rule:

$$\frac{A_1 \quad \dots \quad A_k \quad (A_1 \dots A_k) \rightarrow B}{B} \rightarrow E$$

Provided that for every hypothesis (crossed out formula) on which A_i depends there is a hypothesis on which A_j depends with the same superscript (for any $i, j \leq n$). And provided that the $A_1 \dots A_k$ depend on the same restarts.

If $(A_1 \dots A_k) \rightarrow B$ is true then some of the $A_1 \dots A_k$ may be irrelevant to B and some *are* relevant to B. Suppose that A_1 is irrelevant to B and A_2 is relevant, then the condition on the elimination rule ensures that any irrelevant assumption on which A_1 depends is discharged simultaneously with relevant assumptions on which A_2 depends. Thus, the condition ensures that it is always the case that at least one relevant assumption is discharged (whenever a discharge is made).

It should be no surprise to learn that $(A_1 \dots A_k) \rightarrow B$ is deductively equivalent to $(A_1 \wedge \cdots \wedge A_k) \rightarrow B$.

The way I have set the logic up here, it is more convenient to have separate rules for negation (rather than defining it as $A \rightarrow \bot$).

$$\frac{\begin{array}{c}[\cancel{A}^m] \\ \vdots \\ \bot\end{array}}{\neg A} \neg I \qquad \frac{A \quad \neg A}{\bot} \neg I$$

Conjunction introduction is:

$$\frac{A_1 \quad A_2}{A_1 \wedge A_2} \wedge I$$

Provided that for every hypothesis on which A_i depends there is a hypothesis on which A_j depends with the same superscript (for any $i, j \leq 2$). And provided that the A_i depend on the same restarts.

The side condition ensures that for every application of relevant-restart that is incomplete at A (or B), there is an application of relevant-restart with the same premise that is incomplete at B (or A). In other words, A and B depend on the same applications of relevant-restart.

Conjunction elimination is:

$$\frac{A \wedge B}{A} \wedge E \qquad \frac{A \wedge B}{B} \wedge E$$

Disjunction introduction is

$$\frac{A}{A \vee B} \vee I \qquad \frac{B}{A \vee B} \vee I$$

and disjunction elimination is

$$\frac{A \vee B \quad \begin{matrix}[\cancel{A}^m] \\ \vdots \\ C\end{matrix} \quad \begin{matrix}[\cancel{B}^m] \\ \vdots \\ C\end{matrix}}{C} \vee E(m)$$

Provided that for every hypothesis on which each occurrence of C depends there is a hypothesis on which the other occurrence depends with the same superscript. And provided that both occurrences of C (in the minor premise of the rule application) depend on the same restarts.

The rules for $\perp$:

$$\frac{\perp}{A} \perp E$$

Provided that $\perp$ depends on an incomplete application of restart the premise of which is A

The rules for the universal and existential quantifier are unchanged (as in sections 11.1.1 and 11.1.4, except that the existential elimination rule is this:

$$\frac{\exists x A \quad \begin{matrix}[\cancel{A[x/c]}^m] \\ \vdots \\ B\end{matrix}}{B} \exists E(m)$$

where c is a constant that does not occur in any formulae or (premises of) applications of restart on which B depends except $A[x/c]$, nor in $\exists x A$ nor in B itself.

When referring to this less problematic relevance logic, let us write $\Gamma \vdash A$ when there is a deduction the bottom node of which contains an occurrence of A and the top nodes of which contain only occurrences of members of Γ.

12.4.2 Deduction theorem

The deduction theorem does not hold in this logic, for example $\{A, B\} \vdash A \wedge B$ but $B \nvdash A \rightarrow (A \wedge B)$ is not a theorem.[15] Perhaps we should deny that that really is the deduction theorem, the deduction theorem (for relevance logic) is that

[15] The normal form theorem of the next section may be used to show this, at least for atomic A and B the only deduction in normal form of $A \wedge B$ from A, B is a single application of $\wedge I$, but then neither A nor B can be discharged alone to derive $A \rightarrow (A \wedge B)$.

> if $\Gamma, A \vdash B$ then there is some finite subset $\{A_1 \dots A_n\}$ of Γ such that $\Delta \vdash (A_1 \dots A_n, A) \rightarrow B$ and $\Gamma = \Delta \cup \{A_1 \dots A_n\}$.

To see that this is a theorem note that a deduction of A from Γ is a finite affair. So we may take a deduction that $\Gamma, A \vdash B$ and choose $A_1 \dots A_n$ to be all the undischarged (not crossed out) assumptions of the deduction. We may then cross out all the A_i and superscript them with some integer n (that does not appear elsewhere in the deduction), then we add an application of $\rightarrow I$ with additional label n to the end of the Prawitz tree to get a deduction that $\vdash (A_1 \dots A_n, A) \rightarrow B$. Each A_i is in Γ and so, for the Δ such that $\Gamma = \Delta \cup \{A_1 \dots A_n\}$, we have a deduction that $\Delta \vdash (A_1 \dots A_n, A) \rightarrow B$.

12.4.3 Normalisation

Theorem 12.4.1 Every deduction in the less problematic relevance logic can be reduced to a normal form.

Proof:

I sketch the proof here as it proceeds much as before. For example, suppose $A \vee B$ is introduced and then eliminated:

$$\dfrac{\dfrac{\begin{array}{c}\vdots\\ A\end{array}}{A \vee B} \quad \begin{array}{c}[\cancel{A}^m]\\ \vdots \\ C\end{array} \quad \begin{array}{c}[\cancel{B}^m]\\ \vdots\\ C\end{array}}{C}\ \vee E(m)$$

then we may reduce this section of the Prawitz tree like this:

$$\begin{array}{c}\vdots\\ A\\ \vdots\\ C\end{array}$$

Because the occurrences of C all depend on the same restarts, any application of $\bot I$ or r-restart below C has its side condition met. Furthermore, if one occurrence of C depends on a hypothesis with superscript n, then the does the other. Thus any application of $\rightarrow I$ (or $\neg I$) below C remains legitimate.[16]

[16] The minor premise

$$\begin{array}{c}[\cancel{B}^m]\\ \vdots\\ C\end{array}$$

The cases for the other connectives are largely the same as in previous arguments and I omit them here.

The most problematic cases arise with $\bot$. We can easily reduce this

$$\begin{array}{l}\vdots\\ \underline{A}\ \ \textit{restart}\\ \underline{\bot}\ \ \bot E\\ A\end{array}$$

by replacing it with th initial deduction of A. But more complicated is the reduction of this:

$$\begin{array}{l}\vdots\\ \underline{B}\ \ \textit{restart}\\ \underline{\bot}\ \ \bot E\\ A\end{array}$$

I shall sketch how to reduce these cases. First note that for such a case to arise, B must depend on some prior applications of restart with a premise A. Therefore, the Prawitz tree looks like this:

$$(\dagger)\qquad \begin{array}{l}\vdots\\ \underline{A}\ \ \textit{restart}\\ \bot\\ \vdots\ (w)\\ \underline{B}\ \ \textit{restart}\\ \underline{\bot}\ \ \bot E\\ A\\ \vdots\ (v)\\ B\end{array}$$

Now, suppose that the section Prawitz tree labelled w contains no applications of restart that are completed in the Prawitz tree labelled v. Then we may replace † by this:

$$\begin{array}{l}\vdots\\ A\\ \vdots\ (v)\\ \underline{B}\ \ \textit{restart}\\ \bot\\ \vdots\ (w)\\ B\end{array}$$

is deleted in the reduction step, we have shown that there are no rule applications below C, in the original deduction, that are legitimate only because of this premise and not also because of the other premise.

On the other hand, looking at †, suppose that there are applications of restart in w that are completed in v. Let C be an occurrence of a formula that occurs in v, and completes an application of restart in w, and is such no other formula in v completes an application of restart in w and is higher than C. We may represent this case in this way:

$$(\ddagger)\qquad \begin{array}{cl} \vdots & \\ \dfrac{A}{\bot} & \textit{restart} \\ \vdots & (w_1) \\ \dfrac{C}{\bot} & \textit{restart} \\ \vdots & (w_2) \\ \dfrac{B}{\bot} & \textit{restart} \\ \overline{A} & \bot E \\ \vdots & (v_1) \\ C & \\ \vdots & (v_2) \\ B & \end{array}$$

because of the way we have chosen C, we may replace ‡ with:

$$\begin{array}{cl} \vdots & \\ A & \\ \vdots & (v_1) \\ \dfrac{C}{\bot} & \textit{restart} \\ \vdots & (w_2) \\ \dfrac{B}{\bot} & \textit{restart} \\ \vdots & (w_1) \\ C & \\ \vdots & (v_2) \\ B & \end{array}$$

An essential part of previous normalisation arguments was a proof that any deduction can be reduced to a deduction where no formula is both the conclusion to a restart rule and a premise to an elimination rule (see page 8.3).

In the case of this normalisation theorem we must prove that any deduction may be reduced to one where no formula is both the conclusion of $\perp E$ and the premise of an elimination rule. This can be proved just as in previous normalisation arguments (see around page 8.3).

We obtain normalisation by repeated application of these reduction steps (notice that each reduction step rewrites a part of the deduction so that as a whole the deduction contains one fewer formula that is introduced and then eliminated). ∎

Consider an atomic conclusion C of a deduction from $\{A, \neg A\}$ where A is atomic. Then it may be deduced in normal form and since the conclusion is atomic the last step in an elimination rule. Since the deduction is in normal form C must be deduced via a sequence of elimination rules from assumptions in $\{A, \neg A\}$ but such a sequence can yield only A or $\perp$. Thus it is not the case that everything follows from a contradiction (and similarly from $\perp$) in this logic. More complicated arguments show that $\nvdash A \to (B \to A)$ for any atomic A and B.

It is good that the result that a contradiction does not deduce everything can be shown entirely proof theoretically and without appeal to semantic notions. This is because a semantics of relevance logic seem to be either trivial and unhelpful,[17] or unintelligible and unhelpful.[18]

12.4.4 Some deductions in relevance logic

The new inference rules may be considered as improvements as we now have much more general rules for the logical connectives. The following deductions exemplify their strength. In these deductions I call the restart rule $\perp I$.

$$\dfrac{C \qquad \dfrac{A \quad B}{A \wedge B}\,\wedge I}{(A \wedge B) \wedge C}$$

$$\dfrac{\cancel{A}^{1}}{(A, B) \to A}\,\to I(1)$$

B gets empty discharged, we can do this as one of the antecedents of $(A, B) \to A$ really does appear at a top node of the Prawitz tree.

[17]For example an operational semantics where $A \to B$ means effectively 'there is a deduction in relevance logic of B from A'.

[18]For example the ternary accessibility relation possible world semantics.

$$\dfrac{\neg(A\wedge B) \qquad \dfrac{A \quad B}{A\wedge B}\,\wedge I}{\bot}\,{\to}E$$

$$\dfrac{A\vee B \qquad \dfrac{\cancel{A}^1 \quad \neg A}{\bot}\,{\to}E \qquad \dfrac{\cancel{B}^1 \quad \neg B}{\bot}\,{\to}E}{\bot}\,\vee E(1)$$

$$\dfrac{\dfrac{\dfrac{\dfrac{\cancel{A\wedge B}^1}{A}\,\wedge E \qquad \dfrac{\cancel{A\wedge B}^1}{B}\,\wedge E \qquad \cancel{(A,B)\to C}^2}{C}\,{\to}E}{(A\wedge B)\to C}\,{\to}I(2)}{((A,B)\to C)\to((A\wedge B)\to C)}\,{\to}I(1)$$

$$\dfrac{\dfrac{\dfrac{\cancel{A}^1 \quad \cancel{B}^1}{A\wedge B}\,\wedge I \qquad (A\wedge B)\to C}{C}\,{\to}E}{(A,B)\to C}\,{\to}I(1)$$

$$\dfrac{\dfrac{\dfrac{\dfrac{\cancel{A}^1}{A\vee\neg A}\,\vee I}{\bot}\,\bot I}{\neg A}\,{\to}I(1)}{A\vee\neg A}\,\vee I$$

$$\dfrac{\dfrac{A\vee B \qquad \dfrac{\dfrac{\cancel{C\to E}^1 \quad \dfrac{\cancel{A}^3 \quad \dfrac{\cancel{A\to C\wedge B\to D}^2}{A\to C}\,\wedge E}{C}\,{\to}E}{E}\,{\to}E}{E\vee F}\,\vee I \qquad \dfrac{\dfrac{\cancel{D\to F}^1 \quad \dfrac{\cancel{B}^3 \quad \dfrac{\cancel{A\to C\wedge B\to D}^2}{B\to D}\,\wedge E}{D}\,{\to}E}{F}\,{\to}E}{E\vee F}\,\vee I}{\dfrac{E\vee F}{(A\to C\wedge B\to D)\to E\vee F}\,{\to}I(2)}\,\vee E(3)}{(C\to E, D\to F)\to(A\to C\wedge B\to D)\to E\vee F}\,{\to}I(1)$$

12.4.5 Concluding remarks: the main case against relevance logic

In this chapter I have attempted to argue that the ex falso rule is valid and that its unacceptability is an illusion. However, I find the justification of the ex falso rule not entirely satisfactory and I think we must still respect relevance logic as being potentially the correct formal system.

I hope to have answered in this chapter any proof theoretic worries about relevance logic. The logical consequence relation of the less problematic relevance logic seems to be as a logical consequence relation should be (e.g. it does not have obvious failings such as a lack of conjunction introduction. However, I shall not endorse it as the basis of the correct system of logic.

As I see it, the main arguments against a replacement of classical logic by relevance logic, or against the view that we have been using relevance logic all along are these.

1. There are ways of interpreting negation and $\perp$ so that ex falso is a valid rule, and furthermore the so called paradoxes of strict implication (that a theorem is implied by anything and a contradiction implies everything) are not paradoxes and are in fact true.[19]

2. The model theory obtained from classical first order logic is indispensable. That is, there is something correct about the truth conditional semantics. With the ex falso rule we know that every possible world is characterised by a set of sentences such that for every A either A or $\neg A$ is true at that world but not both. From this the notion of entailment is simple, $A \to B$ just in case $A \vdash B$ (deduction theorem) just in case every possible A-world is also a B-world. This provides us with a good basic account of what the property of truth and relation of validity are (I leave it open whether this account is real or instrumental, either way it is indispensable).

   It is not clear to me what a helpful semantics for any relevance logic looks like. The closest I can come is to suggest a truth conditional semantics for $\perp$, conjunction and disjunction and to read $(A_1 \ldots A_n) \to B$ as 'there is a deduction of B that requires some of the A_i', this is not very helpful.

For these reasons I shall retain classical logic and not advocate this relevance logic. There is excellent reason for applying this relevance logic to cases where the consequence relation is not intended merely to preserve truth (or to preserve truth at all). For example, in the case of reasoning within a fiction we wish our inferences to preserve relevance to the fiction rather

[19] I find it hard to doubt that $B \to (A \to A)$ is not necessarily true, it seems to follow from the fact that $A \to A$ is necessary. Although perhaps what I find hard to doubt is not $B \to (A \to A)$ but in fact $B \supset (A \to A)$, and $B \supset (A \to A)$ is a theorem of relevance logic.

than truth. We can read and play along with an inconsistent story, this is perhaps because many of the inferences we draw are intended only to preserve the spirit of the fiction. Certainly relevance logic, although not the logic from which the logical connectives derive their meaning, is of great use and importance.

Chapter 13

Implication

In this chapter I shall discuss how my analysis of the meanings of logical constants applies to conditionals. So far I have shown how a logic of the material conditional and the strict conditional may be regarded as analytic. I wish now to investigate the extent to which the natural language construction 'if. . . then. . . ' means the same as $\supset$ or $\rightarrow$, and what 'if. . . then. . . ' means when it is does not have the meaning of $\supset$ or $\rightarrow$.

I shall begin by discussing some famous theories of conditionals. First I shall discuss arguments for the theory that any use of 'if. . . then. . . ' may be analysed by $\supset$ (i.e. the theory that all conditionals are material conditionals). I shall reject these arguments and their conclusion, however I think the arguments to support a weaker conclusion that some uses of 'if. . . then. . . ' really are the material conditional. Furthermore I think that similar arguments support the view that some uses of 'if. . . then. . . ' are the strict conditional.

I shall then discuss the conditional probability theory of conditionals (famously advocated by Edgington and Adams). Ultimately I shall reject the theory, mainly on the grounds that it fares no better as a theory of conditionals than a closest-possible-world theory with the additional problem that conditionals do not (under the conditional probability account) express propositions.

Finally I present my account of conditionals which is a variant on Lewis' closest possible world account. I argue that the logic of conditionals is not entirely analytic, some of the rules for 'if. . . then. . . ' do not determine its meaning (the rules relating to the relation of closeness of possible worlds). Nevertheless, there is an analytic core to the logic of the conditional (the rules that determine that it is a conditional). The significance of this for my theory of analyticity is this: the full logic of the conditional is not harmonious, nevertheless there is an important sublogic that is harmonious. That sublogic consists of rules that identify the 'if. . . then. . . ' as a conditional (they are similar to the rules for $\rightarrow$). The remaining rules relate to the ordering of possible worlds by the closeness relation (these rules are learned through experience).

13.1 Arguments that implication is material implication

There are two compelling arguments not discussed enough that the conditional is truth functional is material. The arguments come in a number of forms.

13.1.1 Argument from conjunction

Suppose that we have these intuitive rules for negation:[1]

$$\dfrac{\begin{array}{c}\cancel{A}\\ \vdots\\ \bot\end{array}}{\sim A}\,{\sim}I \qquad \dfrac{A \quad \sim A}{B}\,{\sim}E \qquad \dfrac{\sim\sim A}{A}\,{\sim}DE$$

which are standard rules for classical negation.[2] And these intuitive rules for the conditional

$$\dfrac{\begin{array}{c}\cancel{A}\\ \vdots\\ B\end{array}}{\text{if}\,A\,\text{then}\,B} \qquad \dfrac{A \quad \text{if}\,A\,\text{then}\,B}{B}$$

The argument then proceeds by means of two simple deductions: a de-

[1] The rule ${\sim}I$ is often called *reductio*, the rule ${\sim}DE$ is the infamous *double negation* elimination rule. The rule ${\sim}E$ might be called *contradiction elimination* but has no official name as far as I know.

[2] Usually only ${\sim}DE$ is necessary as $\sim\ldots$ is defined to be $\ulcorner$if $\ldots$ then $\bot\urcorner$. But I do not do this here so that $\sim$ at least appears to be a term in its own right and not defined in terms of the conditional.

duction of ⌜if A then B⌝ from $\sim(A \wedge \sim B)$ and vice versa.

$$\cfrac{\cfrac{\cfrac{\cfrac{\cfrac{\cancel{A}^2 \quad \cancel{\sim B}^1}{A \wedge \sim B} \qquad \sim(A \wedge \sim B)}{\perp}}{\sim\sim B}\,(1)}{B}}{\text{if } A \text{ then } B}\,(2) \qquad\qquad \cfrac{\cfrac{\cfrac{\cancel{A \wedge \sim B}^1}{\sim B} \qquad \cfrac{\text{if } A \text{ then } B \qquad \cfrac{\cancel{A \wedge \sim B}^1}{A}}{B}}{\perp}}{\sim(A \wedge \sim B)}\,(1)$$

Put into words the arguments run like this

- Suppose that one of A and $\sim B$ is false, now suppose also that A, it follows that $\sim B$ is not true (for otherwise A and $\sim B$ are both true), so B is true. Thus if A then B.
- Also, ⌜if A then B⌝ is inconsistent with $A \wedge \sim B$, so each implies the negation of the other. Therefore 'if A then B' implies $\sim(A \wedge \sim B)$.

13.1.2 The argument from disjunction

A similar argument is based on disjunction.

- Suppose $\sim A \vee B$ then if A, B.
- Suppose that if A then B, now there are two cases: either $\sim A$ or A. In the second case B (as we have supposed that if A then B). So the two cases become: $\sim A$ or B, i.e. $\sim A \vee B$.

The conclusion of both arguments is that if we accept modus ponens and conditional proof for the conditional (together with double negation elimination or the restart rule) then the conditional is equivalent to material implication.[3] This then applies to my thesis in which the conditional is defined implicitly by conditional proof and modus ponens.

The arguments are hard to fault, perhaps the first is the more surprising to those not wishing to accept the material conditional as the natural language conditional.

13.1.3 A response

Given the restart rule (which entails double negation elimination), anything that satisfies conditional proof and modus ponens is at least as strong as the material conditional. The conclusion we must draw is that the natural

[3] If we accept that it expresses a proposition, which Edgington and Adams do not.

language conditional does not satisfy conditional proof, at least not in such an unrestricted form.

We have already seen how a restricted form of conditional proof is required when giving the logic of the strict conditional. $A \to B$ may be introduced not on the basis of any deduction of B from A but only on the basis of a deduction that can be made in any context. However the deduction of B from A required to introduce $A \supset B$ may depend on anything that just happens to be true (and may be false in different situations). In particular the deduction may depend on B, if it is true, and so $A \supset B$ can be introduced only because it is true that B.

I think the form of conditional proof we actually use, in general, is restricted in a similar manner, for $\ulcorner$if A then $B\urcorner$ to be true, not just any deduction of B from A must be available. $\ulcorner$if A then $B\urcorner$ is true when B follows from A on the basis of certain appropriate information.[4] I begin a discussion of what information is 'appropriate' means in section 13.5.

Even so, the arguments remain compelling and I think that the material conditional is used. I suggest that the construction:

> if... as well, then...

is a material conditional. For example, the embedded conditional in the latter half of:

> (†) if $A \vee B$ then if $\sim A$ as well, B.

is of the form if... as well, then.... Let us suppose for now that the main connective of † expresses entailment, then † is best analysed by

$$(A \vee B) \to (\sim A \supset C)$$

and is equivalent to $((A \vee B) \wedge \sim A) \to C$. It is easier to see that the 'if as well' construction is material when it is written in a longhand

> if..., in addition to all else that is the case, then...

Now if B is true, then if A is true in addition to all else that is the case (e.g. B) then B is true. Furthermore if $\sim A$ is true, then if A is true in addition to all else then $\perp$ is true, and so (by ex falso) B is true.

It is rare that we use the 'if as well' by itself, usually it occurs embedded in another conditional. A similar construction is 'and if... then ...', which clearly can occur only embedded

[4]See page 203 for the full truth definition for the natural language conditional.

if A and if B then C

it seems to me that the best analysis of such a construction is the material conditional.[5]

13.2 The inescapability of the logical conditionals

Despite the stock of counterexamples to the strict and material conditionals as an analysis of the natural language conditional, the two conditionals exist in natural language and are unavoidable. This follows directly from the arguments above, we may always define such conditional, and such conditionals will always be useful to us as they allow us to express our reasoning and thoughts. Even the most anti-logical text on the subject, loaded with counterexamples to the material and strict analysis of the natural language 'if. . . then. . . ' makes use of a term like 'therefore'. $\ulcorner A$, therefore $B\urcorner$ satisfies all the properties of the strict conditional, it must, for otherwise we would not be able to string an argument together such that the premises are grounds for the conclusion (e.g. we cannot if transitivity of 'therefore' fails).

Indeed, any formulation of the logic of the 'real' conditional makes use of a rule like:

(†) If A logically implies B then, $A \Rightarrow B$

where $\Rightarrow$ is to be the natural language conditional.[6] What is the status of 'logically implies' in such sentence and what is the status of the 'if. . . then. . . ' used explicitly in the sentence? If, in order to grasp the natural language conditional, we are to grasp something like † then we already have, or have the means to define and use the strict conditional, *before* we come to have $\Rightarrow$. Otherwise † the 'if. . . then. . . ' and 'logically implies' of † is $\Rightarrow$ in which case, † is circular if used as a defining axiom for $\Rightarrow$, and adds nothing to the logic except perhaps the theorem $(A \Rightarrow B) \Rightarrow (A \Rightarrow B)$.

A more specific example, it has been argued that the natural language 'if. . . then . . . ' is a judgement best captured by

$P(B|A)$

[5]This also allows us to solve a difficulty for embedded conditionals and modus ponens presented my McGee, see 13.7.2.

[6]For example, see the formulation of the rule *LC* on page 154, of Adams' book *A primer of probability logic*, CSLI, 1998.

understood in the usual probability calculus. We accept 'if. . . then . . . ' when we judge $P(B|A)$ to be high. But to be competent users of the probability calculus we must know, or at least act in such a way that

If A is tautology then $P(A) = 1$

which is an axiom of the probability calculus. In terms of the judgement theory, upon learning that A is a tautology we should be certain of it. But we cannot learn that A is a tautology unless there already exists a framework for deducing tautologies. If such a framework exists then a strict and material conditional is definable. Furthermore, suppose we wish to explain someone why we are certain that $\sim(A \wedge \sim A)$ we do not argue that its probability is 1, for this merely serves to restate that we are certain of it. In fact, we try to show that $\sim(A \wedge \sim A)$ *cannot be false* we show that it is *a tautology*, in doing so we might argue that *if* $A \wedge \sim A$ *then* an inconsistency follows ($\perp$). That use of 'if. . . then. . . ' is the strict conditional and is essential to making the argument understood. The strict conditional is used in natural language.

These arguments apply equally to a material-implication-only theory. I think it is not tenable to hold that the natural language indicative conditional is only ever the material conditional.[7] Even one who loves the material conditional must acknowledge that there is a difference between

(†) If it is sunny then I will not take my umbrella

and

(‡) If it is sunny then it is sunny,

one is necessary and the other is not. The difference can be characterised by using some analogue of a necessity or theoremhood operator such that ‡ is necessary (or a theorem) but † is not. Another way to go, a simpler way, is to accept that, in addition to other conditionals, there is a strict conditional for which ‡ is true and † is not. This way there is no extra notion of necessity to rely on, or even a notion of theoremhood which (in the same way as for a probabilistic conditional above) requires a strict conditional in its formulation. The strict conditional is neither mysterious nor circular as it may be implicitly defined in much the same way as the material conditional.

[7]Talk of 'indicative' rather than 'counterfactual' conditionals comes with a disputable taxonomy of conditional sentences. I use it here to help refer to theories of others who accept the distinction and argue that so called indicative conditionals are truth functional.

13.3 The probability analysis

Before turning to my own analysis of conditionals I want to discuss an account of conditionals that, if true, renders pointless any investigation into the analyticity of conditional sentences. The account is that conditional statements do not express propositions but are express a primitive judgement that is best analysed by a conditional probability (if conditionals do not express propositions then they are neither true nor false and so there is no point in trying to explain why some are true only because of the meaning of 'if... then...'). This probabilistic account of conditionals is accepted by many philosophers, famously Edgington and Adams.

Edgington argues for the thesis when we 'believe' that if A then B we are really making a conditional judgement about B (conditional on A). The degree of conditional-on-A belief we have in B is given by $P(B|A)$ the conditional probability of B given A.

13.3.1 Edgington's argument

Here is, from what I can tell, Edgington's master argument. I change only Edgington's notation in the following quotation (from [**?**, p279-80])

> Two prima facie desirable properties of indicative conditional judgements:
>
> (i) Minimal certainty that $A \vee B$ (ruling out just $A \wedge \sim B$) is enough for certainly that if $\sim A$, B; changing the negation sign, minimal certainly that $\sim A \vee B$ (ruling out just $A \wedge \sim B$) is enough for certainty that if A, B.
>
> (ii) It is not necessarily irrational to disbelieve A and yet disbelieve if A, B.
>
> The truth functional account satisfies (i) but not (ii). Stronger truth conditions may satisfy (ii)... but they cannot satisfy (i); for any stronger truth condition ruling out just $A \wedge \sim B$ leaves open the possibility that 'if A,B' is not true. (p279)

Minimal certainty in $A \vee B$ occurs when $p(A \vee B) = 1$ but $p(A) \neq 1$ and $p(B) \neq 1$. Edgington then goes on to argue that her probability analysis of conditionals satisfies both (i) and (ii) at the cost of conditionals not being propositions.[8] Notice that if we use probabilities as an analysis of certainty then we must be certain of every tautology (and certain that every impossibility is impossible), this is a general problem of many systems. In a context such as this, Edgington has a right to ignore it, as do I.

> We may generalize. Take any proposition, Either it is entailed by $\sim(A \wedge \sim B)$ or it is not. If it is, it will satisfy (i) by not (ii) (when substituted for ⌜if A,B⌝). If it is not, it may satisfy (ii), but it cannot satisfy (i). Conditional judgements interpreted according to the [Conditional Probability] Thesis satisfy both (i) and (ii). So they cannot be interpreted as a belief in any proposition. (p280)

I think the error here is the supposition that a semantic treatment of the conditional (in which conditionals are true or false) will treat certainty merely by attaching probabilities to propositions. For example if we have a possible world analysis of conditionals we might demand that a probability distribution on propositions affects the distribution of the possible worlds.

Take Stalnaker's semantics for the conditional [**?**]. There is an actual world which is a member of a set of possible worlds called the *context set*. The context set is contained in a total domain of all possible worlds. There is a closeness relation on the total domain, such that any world in the context set is closer to the actual world (and also any other world in the context set) than any world outside the context set. ⌜If A then B⌝ is true at a world w when B is true at the closest world to w where A is true. A proposition is true when it is true in the actual world.

Edgington's argument applied here is that we might be minimally certain that $A \vee B$ (i.e. attach appropriate probabilities to A and B and $A \vee B$) in the actual world, but still, in the closest world where $\sim A$ is true B is not true (if A is actually true). I think the error in Edgington's argument lies in assuming that the probabilities of A and B at a world are independent of what worlds are close or distant to that world.

Stalnaker does not write so much about certainty, but he is quite clear on how *knowledge* should be treated in his model when he discusses what the context set is. Stalnaker stresses the need to evaluate conditionals in context...

[8]Though she does allow that we can make assertions out of a conditional probability, if we think $p(B|A)$ is high then we can assert that $p(A \wedge B)$ is significantly greater that the probability of $A \wedge \sim B$.

> The most important element of a context, I suggest, is the common knowledge, or presumed common knowledge and common assumption of the participants of the discourse.(p141)
>
> ⋮
>
> In the possible worlds framework, we can represent this background information by a set of possible worlds – the possible worlds not rules out by the presupposed background information. I will call this set of possible worlds the *context set.* (p142)

When we know A, the context set should be such that A is true at every world in it. If we do not know that A (ignoring unknown logical truths[9]) then $\sim A$ should be true at some world in the context set. Thus, if we have minimal knowledge that $A \vee B$, then we know only (of A and B) that $\sim A$ and $\sim B$ are not both true, and so $A \vee B$ is true throughout the context set and $\sim A$ and $\sim B$ are true at some worlds in the context set. It follows from the conditions on the closeness relation that ⌜if $\sim A$ then B⌝ is true in the actual world (and also any other world in the context set).

Stalnaker does not relate his context set to probabilities, but it is easy to find ways in which he can. For example, we could stipulate that the proper objects of probabilities are worlds. Worlds in the context set all have non-zero probabilities such that the sum of the probabilities of the worlds in the context set is 1. The probability $p(A)$ of a proposition A is the sum of the probabilities of the worlds in the context set at which A is true. We may then define $p(B|A)$ as $p(A \wedge B)/p(A)$. And now we have a stronger than truth conditional account of the conditional, in which it expresses a proposition, which satisfies both of Edgington's (i) and (ii) above.

Certainty does not entail truth, so it does not follow that minimal certainty in $A \vee B$ entails that ⌜if $\sim A$ then B⌝ is true (or even $A \vee B$). However minimal knowledge in $A \vee B$ does, on this upgraded account, entail that if $\sim A$ then B. So if we are minimally certain that $A \vee B$ and we also know that $A \vee B$, then ⌜if $\sim A$ then B⌝ is true, and so we have every right to be certain that if $\sim A$ then B (as we assume we know each individual thing of which we are certain).

[9] As Edgington must as well, for there are logical truths of which we are uncertain yet their probability is 1.

13.3.2 Relevance and similarity

An experienced diver is standing on a diving board over a swimming pool that has been drained of water and has not yet been refilled. There is a sense in which it is true that

(†) If he jumps he will break his neck

but, the probability of him breaking his neck given that he has jumped is very low, after all, he knows not to jump into a pool that has no water in it.

Edgington's thesis can easily account for this, using the phrase such as 'as he is' or 'as things are'. The probability that he will break his neck given that he jumps *as things are*, is very high, but if we do not specify that the jump has to be as things currently are, it becomes very low.

But this phrase 'as things are' if we are not careful should result in the probability of everything (except perhaps some quantum events) being 1 or 0. The answer is that when assessing $\ulcorner$if A then $B\urcorner$ we do not always assess $p(B|A)$ out of context, we may identify some facts Γ relevant to the context and instead assess $p(A|(\bigwedge \Gamma \cup \{B\}))$.

We must do something like this anyway when we come to use general probabilities for assessing likelihoods of individual events or properties of individual things. Gillies ([**?**, p119-125]) uses an example of whether to allow a teenage girl to own a scooter. The crucial question centres around the conditional

if she owns a scooter then she will have an accident

how are we to judge what the probability of *her* having an accident is given that *she* owns a scooter. We could look at (possibly objective) probability of someone crashing given that they own a scooter. But this might not be appropriate, the particular teenager in question might not be represented adequately by the class of all past scooter owners. It is perhaps best to look at the probability of a teenager crashing given they own a scooter. But that might still not be appropriate as, this teenager might be more careful than all other teenagers, and so we should consider the probability of a crash given the owner of the scooter is a careful (slow driving) teenager.[10]

The upshot of this is that for knowledge of general probabilities to help us decide about probabilities of particular things and events,[11] we must find

[10]But notice that this adds further complexity as to judge whether it is appropriate to class this teenager as careful we might have to consider different conditional probabilities (e.g. the probability of a teenager being a careful driver given their age and surroundings).

[11]Which it must, otherwise the theory is impotent to explain how we should accept or reject conditional judgements about particular events or things.

some relevant information (from the context) that determines which general probability we use. Suppose I am debating the conditional probability of Tweety flying away if I open the door to his cage. I cannot judge this at all unless I learn some information about Tweety, e.g. that he is a bird, which seems to make the conditional probability very high. But this might not do justice to the situation (e.g. if Tweety is a penguin).

The conditional probability analysis of conditionals cannot work without a separate theory of identifying an optimal set of relevant facts to use for turning general probabilities into probabilities of individual things and events. Or, perhaps some account of similarity of situation so that when we judge $p(B|A)$ we do so by considering all similar cases where A is true and look at the proportion of them that are B.

Notice a similarity between a theory of optimal sets of relevant facts and a theory of closest possible worlds. The similarity is close to identity, especially if worlds are deemed closer or more distant depending on how many more relevant facts are true at them.

Thus a probability theory of conditionals cannot claim to have an advantage over other theories (e.g. possible world theories) in that it bypasses considerations of similarity of worlds or relevance of information, for it does not. Furthermore, as we have seen, Edgington's master argument does not succeed.

13.3.3 Embedded conditionals

Since $p(B|A)$ is not the probability of a proposition $B|A$ we cannot embed it in another connective. Thus we cannot analyse $\ulcorner \text{if } A \text{ then if } B \text{ then } C \urcorner$ in terms of $p((C|B)|A)$ as the latter has no meaning (or at least, any attempt to give it meaning results in disaster). This means that the probability thesis of conditionals can only account for conditionals with do not occur embedded in a sentence. Proponents of the thesis seek to paraphrase what constructions involving embedded conditionals they can and argue that the remaining ones are unintelligible. This is perhaps an acceptable strategy as complex sentences involving multiple embeddings of conditionals (or even a single embedding of a conditional in the antecedent of another) are rare and hard to understand.

For example $\ulcorner \text{if } A \text{ then if } B \text{ then } C \urcorner$ is commonly paraphrased as $\ulcorner \text{if } A \wedge B \text{ then } C \urcorner$. I have yet to see a reasonable paraphrase of $\ulcorner \exists x \text{ if } Fx \text{ then } Gx \urcorner$ as in 'there is someone who will die if they eat a peanut'.

I do not wish to discuss these worries about Edgington's thesis much, mainly as they can be avoided by altering conditional probability theory of conditionals to one which asserts that $\ulcorner$if A then $B\urcorner$ is a proposition that is true when $p(B|A)$ is sufficiently high. That is, it is an assertion that $p(A \wedge B)$ is significantly greater than $p(A \wedge {\sim}B)$.

I content myself here to state my dissatisfaction with the paraphrases (or lack of paraphrases) presented by proponents of Edgington and Adams' theses. I recommend that the theses be altered as suggested above in order to avoid these difficulties without altering the essence of their positions.

To cast more doubt on the ability to find paraphrases, here is an example of a counterfactual embedded in an indicative conditional. A person dies after eating what turns out to have been a poisoned meal and an autopsy is performed to determine whether the meal was poisoned. A test for a certain poison is carried out and the doctor awaits the result. The doctor may assert:

> If the test yields positive then if he had not eaten the meal he would not have died

I claim that no matter how we add time indexes to this it cannot be paraphrased by a variant of $\ulcorner$if $A \wedge B$ then $C\urcorner$ simply because the test would not have been carried out at all had he not eaten the meal. This kind of example is worrying only to someone trying to apply to a conditional probability thesis both to counterfactuals and indicative conditionals. It seems to me that if the probability thesis is applied to one it should be applied to the other for I find it hard to believe that, given the similarities between the two,[12] one sort of conditional is a proposition whereas the other is not.

13.3.4 Final remark

I do not doubt that a conditional probability has a great deal to do with the content of a conditional. But I see no advantage to accepting analysing conditionals solely in terms of conditional probabilities. Lewis', and other, triviality results show that such an analysis cannot regard conditionals as propositions. This loss, in my opinion, far outweighs the gain, for what gain is there? Edgington's argument above is invalid, and when we come to apply conditional probabilities to particular cases we get just as involved with considerations of similarity and relevance as with the possible world semantics for conditionals.

I shall therefore maintain that conditionals express propositions and reject the conditional probability thesis. However, I shall not regard a theory

[12] For example similarities between counterfactuals and indicative conditionals about the future.

of conditionals as complete unless some relation is made between the truth of a conditional and a conditional probability. That is, there is a sense in which

If I buy a lottery ticket I will not win the lottery

is true, simply because that conditional probability of winning given that I buy a ticket is so low. There is also a sense in which it is false, unless the lottery is biased. On the other hand I think that

If I toss this coin it will land heads

is false, unless the coin biased, as the conditional probability of it landing heads given that I toss it is not high enough (only 0.5).

13.4 Gibbard cases: an argument against the propositional approach

Before turning to my account of conditionals I wish to discuss an argument that conditionals do not, in general, express propositions. The argument is due to Gibbard and is described clearly in [**?**]. I do not think that Gibbard's argument does not support the conclusion that conditionals do not express propositions. However, I do think Gibbard's argument shows that the same conditional expresses different propositions depending on the background knowledge of the speaker.

The premise of Gibbard's argument is that if A is not itself impossible then $\ulcorner \text{if}\, A \,\text{then}\, B \urcorner$ and $\ulcorner \text{if}\, A \,\text{then}\, C \urcorner$ should be incompatible if B and C are incompatible, for nobody could sensibly (or rationally) believe or assert both $\ulcorner \text{if}\, A \,\text{then}\, B \urcorner$ and $\ulcorner \text{if}\, A \,\text{then} \sim B \urcorner$ an also believe A to be consistent. Now, suppose two people Peter and Jane know that three people a, b and c are in a room. Both Peter and Jane know that one of a, b, c will remain in the room and the other two will leave. Looking into the room Peter sees a leave and Jane sees c leave. Peter concludes 'if b leaves then c is alone in the room', but Jane concludes 'if b leaves then a is alone in the room'. The problem is then that these two conditionals seem both true, and yet by the argument above they should be incompatible ('b leaves the room' is not contradiction and 'a is alone in the room' is incompatible with 'c is alone in the room'). There is nothing unique to Jane's or Peter's position that allows us to say that e.g. Peter's conditional is true but Jane's is not. Edgington concludes from this that both conditionals are not true, nor are they both false, the

Gibbard case is yet another example of how a conditional does not express a proposition.

A proponent of the thesis that the natural language conditional is the material conditional has no problem with the Gibbard case itself for he allows that $A \supset B$ and $A \supset {\sim}B$ may both be sensibly believed (even by the same person) if A is false, so much the worse for the proponent of the material conditional.

I shall respond to the Gibbard cases first by attacking the relation between the initial claim about incompatible conditionals and the argument itself. Consider the following analogous argument:

> The two expressions 'I am F' and 'I am G' are incompatible if 'is F' and 'is G' are incompatible predicates, the reason for this is someone could not sensibly (or rationally) assert or believe both. Now, any two people could easily find incompatible predicates F and G such that one truly asserts 'I am F' and the other truly asserts 'I am G' (e.g. 'I am the candidate' vs. 'I am the examiner').

What does this tell us about the personal pronoun? It tells us not that it is not a referring expression, but that its primary semantic contribution is its *character* (a function from information states to contents) rather than a unique content. Although no *one* person could reasonably assert both 'I am F' and 'I am not F' there is nothing wrong with two different people asserting them truly.

The moral of the story is that the content of the conditional depends on the context. I conclude from this that the semantic value of a conditional is its character: a function from contexts states to contents. As far as I can see the only relevant feature of the context in the Gibbard cases is the information state of its utterer. Therefore I suggest that the semantic value of a conditional is a function from information states for contents.

Therefore, sometimes two people who assert incompatible conditionals are not really in dispute. In the example above, Peter asserts that if b leaves then c is alone in the room and Jane asserts that if b leaves then a is alone. Naturally, in such a case, Peter and Jane have nothing to argue about as it is clear that they are speaking from different knowledge states. Both can then conclude that b is alone in the room.

Two people who argue over whether $\ulcorner$if A then $B\urcorner$ and are in genuine dispute, do so assuming, tacitly, that each expresses the same proposition by that conditional. Or the dispute could arise when each believes they have all relevant information about the consequences of A.

For example, suppose Peter knows that $A \vee B$ but is undecided about each of the disjuncts, and Jane knows that A. They then dispute whether ⌜if not A then B⌝ is true. Given Jane's belief state she interprets it more like a counterfactual, but Peter does not and a disagreement arises. In this simple case we should expect the disputers to realise quite quickly what is happening, and come to some agreement as to which conditional proposition to debate.

13.5 Variable strict conditionals, my analysis

13.5.1 Dependent strict conditionals

The strict conditional is far too strict to do for the natural language conditional. $A \to B$ is true when and inference of B from A is truth preserving independently of any other truths that are not analytic (logical truths). Clearly we do not intend to say anything so strong with our normal use of the conditional (neither do we intend to say $(A \wedge L) \to B$ where L is a conjunction of logical laws).

Further the logic of the strict conditional is wrong. The general logic of the natural language conditional does not support an inference like

> if A then C, therefore if $A \wedge B$ then C[13]

counterexamples to this are common, for example:

> (†) if I go to central China I will have trouble communicating with the locals.

is true as I have no Chinese and, I am sure, central Chinese have little English. But

> (‡) if I go to central China and everyone there speaks English then I will have trouble communicating with the locals.

is false.

There are many more counterexamples, the most convincing of which run on similar lines. Notice that it is perhaps plausible that I should go to central China, but it is considerably less plausible that I go to China and that everyone there speaks English.

Some (e.g. Lowe [?]) argue that that such examples do not show that the logic of the natural language conditional is any different from that of

[13] The 'therefore' denotes entailment and is, therefore, the strict conditional $\to$.

the strict conditional. The response to the purported counterexamples is that the conditional is highly context sensitive, † does entail ‡ but not if the context of ‡ is different from that of †. But it is hard to see how the context of the two conditionals could have changed, save by virtue of the different antecedent. There are cases of an utterance altering its own context of utterance (e.g. a sarcastic remark), but as far as I can see this happens only because of some property external to the content of the utterance itself (e.g. intonation). In the case of † and ‡ the difference is entirely internal. Thus, even if the conditional † and ‡ differs only in its context, this context change is brought about by the conditional itself and the logic we give of the conditional should reflect this.

On the other hand there is an argument to be made that it should not. Inferences to which we find counterexamples like that of above are commonly made for example

$$\frac{\begin{array}{l}\text{if } A \text{ then } B\\ \text{if } B \text{ then } C\end{array}}{\text{if } A \text{ then } C}$$

has counterexamples along the lines of the one above, and yet seems a natural inference to make, also the antecedent strengthening inference for which we have a counterexample above is commonly made. A natural answer to this is that in these inferences all the conditionals are in the same context. If we alter the logic of the conditionals we must indicate how some acceptable instances of generally invalid conditional reasoning are valid. I find it unacceptable to argue that people are so often wrong about the way they reason.[14]

The most famous analysis of the conditional (Ramsey) shows us how to reconcile this dilemma. I assume (as does nearly everyone else it seems) that we judge that $\ulcorner$if A then $B\urcorner$ on the following basis

> Assume A, make appropriate adjustments for consistency, and then infer that B

[14] Usual modifications of conditional logic support this inference

$$\frac{\begin{array}{l}\text{if } A \text{ then } B\\ \text{if } A \wedge B \text{ then } C\end{array}}{\text{if } A \text{ then } C}$$

it is not implausible that some ellipsis or some pragmatic operation occurs so that reasoning with successive conditionals is really of this form, when it is on the surface an instance of the invalid principle of the transitivity of the conditional. Perhaps other reasoning such as contraposition and the strengthening of the antecedent, which are sometimes valid, can be handled in this way, I shall not investigate how.

that is, if B is judged true conditionally on the assumption that A, with appropriate modifications to our beliefs, then $\ulcorner \text{if } A \text{ then } B \urcorner$ is true. There are at least three ways of interpreting what an 'appropriate adjustment for consistency' is:

1. Evaluate at the closest/most similar possible worlds where A is true.

2. Suspend some beliefs and perhaps assume some extra ones, as is most appropriate, to allow the compatibility of A.

3. Make a conditional probability judgement.

13.5.2 My semantics for the conditional

I wish to avoid a discussion of possible worlds, and I do not wish to get involved with the conditional probability thesis of conditionals. Furthermore, neither fits well with my general analysis of implicit definitions. Thus the analysis I shall adopt is the second, in terms of belief revisions. This suggests the following truth definition for $\ulcorner \text{if } A \text{ then } B \urcorner$:

> $\ulcorner \text{if } A \text{ then } B \urcorner$ is true when either (1) the inference from A to B is valid when dependent on any of the most appropriate revisions of the speakers *knowledge* compatible with A, or (2) A and B are both true.[15]

with this truth definition I give up, to some extent, the objectivity of the conditional, as its truth is dependent on what our background knowledge is. In the discussion of Gibbard cases (13.4) we have an independent argument this is the case. However I think that there are objectively most appropriate revisions of a body of knowledge to make A compatible.

We can use this truth definition to resolve the dilemma above. Let $B \rightarrow_A C$ be true just in case the inference from B to C is valid given the most appropriate revision of the background beliefs to make A consistent. Then, as we shall see

$$\frac{\begin{array}{c} A \rightarrow_A B \\ B \rightarrow_B C \end{array}}{A \rightarrow_A C}$$

is not valid, however

[15]Even if we have no idea what is going on, and we say $\ulcorner \text{if } A \text{ then } B \urcorner$ and it turns out by some fluke that A and B are (or were) true, then we were right although lucky.

$$\frac{\begin{array}{c}A \rightarrow_A B\\ B \rightarrow_A C\end{array}}{A \rightarrow_A C}$$

is valid. In general ⌜if A then B⌝ is to be analysed as $A{\rightarrow_A}B$, and sometimes, once we start using $\rightarrow_A$, we continue using $\rightarrow_A$ for further conditionals. So for example, the following case of contraposition is valid:

$$\frac{A \rightarrow_A B}{\sim B \rightarrow_A \sim A}$$

but this

$$\frac{A \rightarrow_A B}{\sim B \rightarrow_{\sim B} \sim A}$$

is not valid. I claim that many of our supposedly invalid uses of conditional principles are in fact valid, but where we are intending our conditionals to be evaluated in the same (revised) belief state (i.e. the same context).

13.5.3 The most appropriate belief revision

On Quine's analogy of our web of belief, the most appropriate revision of our beliefs to make A consistent revises as few inner beliefs as possible. The more inner beliefs we revise the more radical the revision, the most appropriate revisions for determining conditionals are the least radical. Say the more radical the revision required to make A consistent with our beliefs, the greater its *degree of incompatibility*.

It would be wrong to argue that knowledge of actual degrees of incompatibility, or even the meaning of a 'degree' of incompatibility, is an a priori matter. A judgements about what update of our beliefs is most appropriate, in the light of new information, is affected by our experiences. Thus we should expect that many truths regarding the comparative compatibility of two sentences to be a posteriori. For example, I judge that it is a greater departure from reality that my house will collapse in five minutes than it is that I will scratch my ear in five minutes. Whatever this knowledge is (e.g. knowledge of probabilities or knowledge of belief updating) it is mostly based on experience and not on the meanings of words.

Judgements about degrees of incompatibility come as the result of experience. Our experiences tell us which of our beliefs are better suited to be placed further inside our web of belief. For example, my experience of the weather in England tells me to maintain my belief that it rains in the winter almost to the point of it being a physical law.

I see no reason why, given a set of background beliefs that are incompatible with A, there should not be some *objectively* most appropriate update(s) to it to make A consistent. I assume furthermore that there is a fact of the matter about what we do or do not know.

Also, I leave open the possibility that two difference sets of beliefs should not update most appropriately to the same revised set of beliefs. I think that commonly they do, my knowledge of the European union is different from many others, yet we may still have similar visions of doom when we consider what will happen if Britain joins the single currency.

13.5.4 A theory of the most appropriate belief revision

Lewis famously suggested some conditions on making appropriate updates to accommodate new information to apply to his theory of counterfactuals. Lewis bases his work on treating possible worlds as concrete objects and gives a brief theory of the closeness (or similarity) relation on worlds. Although I am not happy with his theory of possible worlds, his theory of similarity can be relabelled for my purposes as a theory of degrees of compatibility.[16] So, following Lewis, when determining which possible states of affairs (worlds) in which A is true are closest to a possible state of affairs w:

1. It is of the first importance to avoid big, widespread, diverse violations of law.

2. It is of the second importance to maximize the spatio-temporal region throughout which perfect match of particular fact prevails.

3. It is of the third importance to avoid even small, localised, or simple violation of law.

4. It is of little or no importance to secure approximate similarity of particular fact, even in matters which concern us greatly. [**?**]

Put in terms of knowledge revision, the most appropriate revisions of knowledge that make A compatible are the ones that most adhere to these rules above. So a revision that revises a law is more radical (and less appropriate) than a revision that avoids revising a law. Put in terms of a web of belief, the more important it is to avoid revising something, the closer it is to the centre of the web of belief.

[16] Or a theory of the ordering on the web of belief.

There are difficulties with these conditions when applied to counterfactuals, in particular the difficulty lies with the final condition. Some example suggest that it should be of little importance, others suggest that it should be of no importance to secure similarly if particular fact. For example the truth of

> (†) If I had woken up one or two hours earlier the sun would have risen at the time it actually did.

suggests that when (temporarily) updating my beliefs to accommodate the new 'information' that I woke up earlier I maintain the fact relating to the time the sun rose. However the falsity of

> If I had woken up one or two hours earlier then I would have woken up one hour earlier.

suggests that I not should try to make the time I woke up as similar as possible to the actual time I woke up.[17] It seems that some matters of fact should be retained and others should not when carrying out an update. It is easy enough to see which matters of fact should be retained, matters of fact that are causally independent of the A should be retained when trying to update by A, but other matters of fact should not. Thus, the rising of the sun is causally independent of what time I wake up so when updating my beliefs to include the antecedent of † I retain the time at which the sun rose.

I think it is a mistake to attempt to give a purely logical analysis of judgements of degrees of incompatibility, experience plays a crucial role in such judgements. Looking at Lewis' conditions, aside from the first condition (referring to laws which are presumably determined a posteriori) a posteriori knowledge plays too little a role in the update. The kind of a posteriori knowledge I have in mind is knowledge of causal dependencies.

The conditions I suggest for updating some beliefs with A are

Strong updating

1. It is of the first importance to avoid violations of causal or law.

---

[17]It seems that on Lewis' recommendations the counterfactual

> If I had woken up one or two hours earlier then I would have woken up one hour earlier.

is true.

2. It is of the second importance to maintain perfect match of particular fact causally independent of A.

3. It is of the third importance to avoid violations of causal or probabilistic dispositions.

4. It is of the fourth importance to maintain perfect match of particular fact probabilistically independent of A and independent of any causal dispositions surrounding A.

5. It is of no importance to secure approximate similarity of particular fact, even in matters which concern us greatly.

The causal laws are the laws given by physical sciences, a causal disposition is a more general rule of thumb about causal relations. For example, many of the facts we know of psychology are so general (with all kinds of hidden ceteris paribus clauses) that they do not enjoy the same status as, say, a law of physics. Furthermore, I know that in general it does not rain heavily for five days on the trot in the English summer, this is not a law, but it is something I know and is a fact about the dispositions of the weather in England. Thus, when making an update of A to our beliefs

1. First we try to retain causal theories like the laws of physics.

2. Then we keep all (known) beliefs that these causal theories say are independent of A.

3. Then we try to retain the more general (or 'folk') causal theories, like non-scientific theories about the dispositions and behaviours of particular objects (like my beliefs about someone's personality).

4. Then we keep all (known) beliefs that these folk causal theories say are independent of A

Perhaps we can eliminate talk of causal relations or dispositions by replacing it with talk of probabilistic dependencies. But I doubt it, the failure of probabilistic theories of causation suggests to me that, at least here, we must take knowledge of causation as irreducible. In doing so I give up hope of obtaining a reductive analysis of causation in terms of conditionals, but suggest that the two are developed together. As we gain in experience our causal theories become more subtle and advanced as do our ability to make conditional judgements.

13.5.5 Indicative and counterfactual conditionals

The distinction between indicative and counterfactual conditionals is well known and well debated. Examples such as these

1. If there is an elephant in front of me then there is something wrong with my eyes.

2. If there were (now) an elephant in front of me then there would (now) be something wrong with my eyes.

leave little doubt that there are two different sorts of conditionals at work, or maybe, two different types of context for the same conditional.[18] There are equally well known problems with discerning these conditionals solely in terms of the surface grammar, some conditionals in the indicative mood, e.g. some conditionals about the future, behave like counterfactual conditionals.

If we understand conditionals in terms of inferring the consequent from updating our beliefs with the antecedent then we can differentiate the two sorts of conditional in terms of the update procedure. Looking at the elephant example, in the first conditional I retain the facts of our visual experience when adding the condition that there is an elephant in front me. In the second I do not, if there were an elephant in front of me I would have a visual experience of it. This suggests that there is a second way of determining degrees of incompatibility:

Weak updating

1. It is first importance to secure approximate similarity of particular fact, especially in matters which concern us greatly.

2. It is of the second importance to avoid violations of causal or law.

3. It is of the third importance to maintain perfect match of particular fact causally independent of A.

4. It is of the little importance to avoid violations of causal or probabilistic dispositions.

So I propose that there are two sorts of conditional. One where the degrees of incompatibility are based on strong updating and one where the

[18] This difference cannot be put down to a difference of tense as both conditionals are in the present tense.

degrees of incompatibility is based on weak updating. In terms of the familiar indicative/counterfactual conditional distinction, weak updating gives us indicative conditionals whereas strong updating gives us counterfactual conditionals.[19]

For example, suppose an experienced skydiver is standing by the airlock at 30,000 feet, he is not wearing a parachute. Consider the following conditional:

(†) If he jumps he will fall all the way and die

this conditional could be taken in one of two ways. Since the skydiver is ill-equipped for a jump, it is true that if he jumps he will fall and die.[20] But on the other hand there is a reading where this conditional is false, the man is experienced, he is not going to jump out unless he puts his parachute on, if he jumps he will be wearing a parachute and will not fall and die.[21] We can get a similar ambiguity out of the conditional:

(‡) If he were to jump he would fall all the way and die.

The difference between the two readings is that in one reading we use weak updating and in the second reading we use strong updating.

The result is similar with conditionals about the past. For example compare

> If I was not using a computer five minutes ago then I was hallucinating

with

> If I had not been using a computer five minutes ago then I would have been hallucinating

Given what I saw five minutes ago (a computer) the first is a far better candidate for being true than the second. To evaluate the first we retain the information about my visual experience, this is weak updating; to evaluate the second we do not, this is strong updating.

[19] It falls out of this that where no update is required the indicative and counterfactual conditionals agree.

[20] The conditional is genuinely used in this way, someone might ask me if the skydiver is wearing a parachute and I could respond to you in the negative by saying, truthfully, that †.

[21] Since the diver is experienced we know he will not jump, so we must do some updating in order to accommodate that he will jump, the update is that he is wearing or will put on a parachute.

13.6 A logic of conditionals

Basically, I shall analyse $\ulcorner$if A then $B\urcorner$ as $A \to_A B$, where $\to_A$ is a weakest strict conditional that does not rule out A (if there is such a conditional).

13.6.1 Syntax

In addition to the conditional $\to$ there, for each formula of the language A there is a conditional $\to_A$. For simplicity I shall define the logic in terms of $\Box$ and $\Box_A$ setting $B \to_A C$ to be $\Box_A(A \supset B)$.

ST $\Box_A B$ and $\sim\Box_A B$ are *medium*. $\Box B$ is both medium and strong.[22]

The introduction and elimination rules for $\Box$ are as usual

$$\frac{\Box B}{B}\ \Box E \qquad \frac{C_1 \ldots C_n \qquad \begin{array}{c}\cancel{C_1}^m \ldots \cancel{C_n}^m\\ \vdots \\ B\end{array}}{\Box B}\ \Box I(m)$$

provided that the C_i are all strong, and the inference of B depends on no assumptions other than $C_1 \ldots C_n$ nor on any weak rule applications.

Let $\Diamond_A B$ be defined to be $\sim\Box_A \sim B$. The rules that define $\Box_A$ as a form of necessity operator and hence $\to_A$ as a strict conditional are these:[23]

$$\frac{\Box_A B_1 \ldots \Box_A B_n \quad \Box(B_1 \wedge \cdots \wedge B_n \supset C)}{\Box_A C}\ \Box_A P \qquad \frac{\Box_A B}{B}\ \Box_A E$$

the rule relating the $\Box_A$ to $\Box$ is this:

$$\frac{\Box B}{\Box_A B}\ \Box_A S$$

The rule $\Box_A S$ can be removed by allowing the $\Box I$ also to introduce $\Box_A$ (I think this is the true rule but I use $\Box_A S$ for simplicity of presentation and

[22] For ease we can add that $\sim\Box B$ is medium, but it is not necessary to do so.

[23] $\Box_A P$ is a form of conditional proof, it may be rewritten like this:

$$\frac{\Box_A \quad C_1 \ldots C_m \quad B_1 \ldots \Box_A B_n \qquad \begin{array}{c}\cancel{B_1}^m \ldots \cancel{B_n}^m \cancel{C_1}^m \ldots \cancel{C_m}^m\\ \vdots \\ D\end{array}}{\Box_A D}\ \Box_A P(m)$$

provided that the C_i are all strong, and the inference of B depends on no assumptions other than $B_1 \ldots B_m C_1 \ldots C_m$ nor on any weak rule applications.

proof below). It is then a simple matter to show that this stronger $\Box I$ (which may also be called $\Box_A I$) together with $\Box E$, $\Box_A P$ and $\Box_A E$ normalises. Consequently any sentence deducible with just these rules is analytic.

The rules relating $\Box_A$ to $\Box_B$ are these:

$$\frac{\Box(B \supset A) \quad \Box_B C}{\Box_A C}\ \Box_A O \qquad \frac{\Diamond_B A \quad \Box_B C}{\Box_A C}\ \Box_A T \qquad \frac{\Box_A \sim A \quad \Box_A C}{\Box_B C}\ \Box_A F$$

and finally there is this rule:

$$\frac{A \quad B}{\Box_A B}\ \Box_A M \ \text{provided that } B \text{ is medium}$$

As I have presented these rules I have given up all hope of proving an easy normalisation theorem for the logic as a whole. But normalisation is not necessary for as I have argued above the logic of these variable strict conditionals is not entirely a priori or analytic. The essential ingredient to the logic is developed over time for it take time to develop an ability to judge degrees of incompatibility.

Deductions

A useful derived rule is this

$$\frac{\Box_\perp(B \supset A) \quad \Box_B C}{\Box_A C}$$

By the rules $\Box_A T$ and $\Box_A F$, we have that $\{{\sim}\Box_A C, \Box_B C\} \vdash \{\Box_B{\sim}A, \Diamond_B B\}$. Furthermore, since $\Box(\perp \supset B)$ is easily derivable, we have that $\Box_\perp(B \supset A) \vdash \Box_B(B \supset A)$ using the rule $\Box_A P$. And so

$$\{{\sim}\Box_A C, \Box_B C, \Box_\perp(B \supset A)\} \vdash \{\Box_B \sim A, \Diamond_B B, \Box_B(B \supset A)\}$$

using $\Box_A P$ we can deduce that $\{\Box_B{\sim}A, \Diamond_B B, \Box_B(B \supset A)\} \vdash \perp$ and so that $\{\Box_B C, \Box_\perp(B \supset A)\} \vdash \Box_A C$.

Also, since $A \wedge B$ entails A, we can deduce $\Box_A(B \supset C)$ from $\Box_{A\wedge B}(B \supset C)$ and so

$$\{\Box_A(A \supset B), \Box_{A\wedge B}(B \supset C)\} \vdash \Box_A(A \supset C)$$

using $\Box_A P$.

Furthermore by $\Box_A T$ we get that $\{\Diamond_A(A \wedge B), \Box_A(A \supset C)\} \vdash \Box_{A\wedge B}(A \supset C)$ which by $\Box_A P$

$$\{\Diamond_A(A \wedge B), \Box_A(A \supset C)\} \vdash \Box_{A\wedge B}((A \wedge B) \supset C)$$

The upshot of this is that if $\Diamond_A(A \wedge B)$ (i.e. $\sim A \rightarrow_A \sim B$) then

$$A \rightarrow_A (B \supset C) \equiv (A \wedge B) \rightarrow_{A \wedge B} C$$

13.6.2 Semantics

I shall obtain a semantics similar to that of Lewis. A model consists of a set of worlds W and a system of spheres S_w for each $w \in W$ and a reflexive and transitive accessibility relation R on the worlds. Each sphere $S \in S_w$ is a set of worlds of W and for any two spheres S and S' in S_w, either $S \subseteq S'$ or $S' \subseteq S$. S_w contains a smallest sphere, i.e. S_w has a member that is a subset of all members of S_w. If w' is in a sphere in S_w (the system of spheres around w) then wRw'. Every sphere in S_w contains w. Further, if S is the smallest sphere in S_w and $w' \in S$, then $S_{w'} = S_w$, and $w'Rw''$ iff wRw'. Finally, for every formula A there is a sphere $S_w^A \in S_w$ which is either the smallest sphere that contains a world where A is true, or, if there is no such sphere, is $\bigcup S_w$.[24]

- A formula is medium when whenever it is true at a world w it is true at all worlds in the smallest sphere in w. A formula is strong when it is true at all w' s.t. wRw' whenever it is true at w, for any w.
- $\Box A$ is true at w when A is true at all accessible w'.
- $\Box_A B$ is true at w when B is true throughout the sphere $S_w^A \in S_w$
- The other connectives receive their usual interpretations.

Soundness

The argument that the rules for $\Box$ are sound is similar to that of 10.2.2 (with a similar treatment of restart and PIP). Furthermore since any two members w' and w'' of the smallest sphere in S_w agree that their system of spheres is S_w, the same formulae of the form $\Box_A B$ and $\Box B$ (as $w' \in S_w^C$ implies wRw' for any C) are true at w' and w'' as at w. Thus the condition that any $\Box_A B$ is medium is sound.

[24] In other words, if A is false at every world in every sphere around w then S_w^A contains every world in every sphere around w.

$\Box_A P$ If $\Box_A B_1 \ldots \Box_A B_n$ and $\Box(B_1 \wedge \cdots \wedge B_n \supset C)$ are true at w then C is true at every w' s.t. wRw' where $B_1 \wedge \cdots \wedge B_n$ is true. But every world in S_w^A is such a world and so $\Box_A C$ is true at w.

$\Box_A E$ This is sound as every sphere in S_w contains w.

$\Box_A S$ Since $w' \in S_w^A$ entails that wRw', it follows that if B is true at every accessible w' from w it is true at every world in S_w^A.

$\Box_A O$ If $\Box(B \supset A)$ is true at w then A is true at every world in $\bigcup S_w$ where B is true. Therefore, since the spheres in S_w, are nested $S_w^A \subseteq S_w^B$, and so if $\Box_B C$ is true at w, so is $\Box_A C$.

$\Box_A T$ If $\Diamond_B A$ is true at w then A is true at some world in S_w^B, and since the spheres are nested and S_w^A is the smallest sphere containing a world where A is true, $S_w^A \subseteq S_w^B$.

$\Box_A F$ If $\Box A \sim A$ then there is no sphere in S_w containing a world where A is true. Thus S_w^A is $\bigcup S_w$ and so contains every sphere in S_w.

$\Box_A M$ If A is true at w then S_w^A is the smallest sphere in S_w, so any two members of S_w^A agree on their system of spheres and so if B is true at w and is medium then B is true at all $w' \in S_w^A$ and so $\Box_A B$ is true at w.

Completeness

We now show that the logic is complete for this semantics. For the canonical model take all maximal consistent sets M as the set of worlds. Set

$$mRm' \text{ iff } A \in m' \text{ for all } A \text{ s.t. } \Box A \in m$$

furthermore set

$$m' \in S_m^A \text{ iff } B \in m' \text{ for all } B \text{ s.t. } \Box_A B \in m$$

and let S_m be the set of all S_m^A for every A.

Take truth at a world to be membership in it. We must verify that this meets the correct conditions of the model.

The proof that the definition of R yields a reflexive transitive relation with the correct truth conditions for $\Box$ is similar to that of 10.2.2.

If $\Box_A B \in m$ then by definition $B \in m'$ for every $m' \in S_m^A$. Conversely $A \in m'$ for all $m' \in S_m^A$. Then since M contains all maximal consistent sets and the m' are exactly the sets containing $\{C : \Box_A C \in m\}$ it follows that $\{C : \Box_A C \in m\} \vdash A$, Since the deduction of this must be finite we can find $C_1 \dots C_n$ from $\{C : \Box_A C \in m\}$ such that $\{C_1 \dots C_n\} \vdash B$ and so $\vdash \Box(C_1 \wedge \cdots \wedge C_n \supset B)$, then by the rule $\Box_A P$ we have that $\{\Box C : \Box C \in m\} \vdash \Box_A B$ and thence $\Box_A B \in m$.

So the truth conditions for $\Box_A$ are verified, now we must show that the conditions on the structure of S_m hold.

Suppose $A \in m'$ for no world in $\bigcup S_m$, then $\Box_A \sim A \in m$. Suppose further that $m' \in S_w^B$, now if $\Box_A C \in m$ then, by $\Box_A F$, $\Box_B C \in m$ and so $B \in m'$. Thus if A is true at no world in $\bigcup S_m$, $\bigcup S_m \subseteq S_m^A$. Furthermore $S_w^A \subseteq \bigcup S_m$ by definition. Therefore, if A is true at no world in $\bigcup S_m$, $\bigcup S_m = S_m^A$.

If $m' \in S_w^A$ then $C \in m'$ for every $\Box_A C \in m$. But then if $\Box C \in m$ then $\Box_A C \in m$ (by $\Box_A S$) and so $A \in m'$. Therefore, if $m' \in S_m^A$ then mRm'.

If $\Box_C B \in m$ then $B \in m$ by $\Box_C E$, and so $m \in S_m^C$ for every C. Thus m is in every sphere in S_m.

Suppose S_m^A and S_m^B are in S_m and suppose further that neither $S_m^A \subseteq S_m^B$ nor $S_m^B \subseteq S_m^A$.

1. If $\Box_A \sim A \in m$ then $S_m^B \subseteq \bigcup S_w = S_w^A$ (similarly for $\Box_B \sim B \in m$), so we may suppose that $\Diamond_A A$ and $\Diamond_B B$ are both in m.

2. If $\Diamond_B A \in m$ then $\Box_A C \in m$ whenever $\Box_B C \in m$ (by $\Box_A T$), and so $S_w^A \subseteq S_w^B$ (similarly for $\Diamond_A B \in m$). Therefore we may suppose that $\Box_A \sim B$ and $\Box_B \sim A$ are both in m.

Since $\Box(A \supset A \vee B)$ and $\Box(B \supset B \vee C)$, by $\Box_A 0$ and (2), $\Box_{A \vee B} \sim A$ and $\Box_{A \vee B} \sim B$ are in m. But then using $\Box_A P$ we get that $\Box_{A \vee B} \sim (A \vee B) \in m$. But then $A \vee B \in m'$ for no m' in $\bigcup S_w$, but this is in contradiction with (1). Thus for any two S_m^A and S_m^B in S_m, either $S_m^A \subseteq S_m^B$ or $S_m^B \subseteq S_m^A$.

Since $\top$ is in every $m \in M$, $S_m^\top$ is always the smallest sphere in S_m. Suppose now that and that $m' \in S_m$. Since $\Box_B C$ is medium and $\top \in m$, if $\Box_B C \in m$ then $\Box_\top \Box_B C \in m$ by $\Box_A O$. Thence $\Box_B C \in m'$. Further, if $\Box_B C \in m$ then since $\sim\Box_B C$ is also medium by similar reasoning it follows that $b_B C \notin m'$. Therefore any $\Box_B C \in m$ iff $\Box_B C \in m'$. So $m'' \in S_m^B$ iff $C \in m''$ for every $\Box_B C \in m$ iff $C \in m''$ for every $\Box_B C \in m'$ iff $m'' \in S_{m'}^B$. And so $S_m = S_{m'}$.

Thus if $\Gamma \vDash A$ then in the canonical model outlined above A is true at every world where Γ is true. Since the worlds are maximal consistent sets it follows that $\Gamma \cup \{\sim A\}$ is inconsistent and so that $\Gamma \vdash A$ (by restart or PIP). This concludes the completeness argument.

What we have shown is that the logic is complete for models where there is a total ordering on possible worlds and where the closest worlds to w agree with w on the ordering of the possible worlds. This entails that what Lewis calls the inner modality, here the operator $\Box_\top$, is an $S5$ operator. The outer modality $\Box_\bot$ yields the modal logic KT. And what I call the full modality $\Box$ is $S4$.

13.6.3 The various conditionals and probabilities

$\ulcorner$if A then $B\urcorner$ could mean a number of things, and I think it varies depending on the context and the speaker.

1. $(A \wedge B) \vee (A \rightarrow_A B)$
2. $(A \wedge B) \vee (\Diamond_A A \wedge (A \rightarrow_A B))$

and each one of the above can be interpreted in two different ways: in terms of weak updating or strong updating.

In general, I think people expect conditionals with impossible antecedents to be false, at least when not used in the context of strict mathematical or logical reasoning. So the strict conditional is I think best analysed by 2.[25]

The logic of $\Box_\top$ (the 'inner modality') is that of $S5$, the sphere of $\Box_\top$ (the smallest sphere) will do nicely as the analogue of Stanlaker's context set. $\Diamond_\top A$ is true if A is compatible with what we is know.[26] We can then add a probability distribution to it.[27]

[25]Though it may be that some considerations of Lowe are correct and that the best analysis is this:

$$(A \wedge B) \vee (\Diamond_A A \wedge (A \rightarrow_A B)) \vee \Box_\bot B$$

[26]This does not mean that $\Box_\top A$ is the analysis of 'it is known that...', there is a difference between $\sim\!A$ being incompatible with what I know and my knowing A.

[27]For example, to each system of spheres S_w we can put attach a probability distribution p_w on the members of $S_w^\top$. We may then extend it to apply to all formulae by setting $p_w(A) = \Sigma_{w' \in S_w^\top} p_w(w')$. Further let $B\|A$ be true at a world w whenever $p_w(B|A) \geq 0.95d$ or some other significant value. Notice that $B\|A$ is true at w iff it is true at each world $w' \in S_w^\top$.

We may then differentiate between two conditionals:

1. If A then probably B
2. If A then definitely B

The second of these is simply $(A \wedge B) \vee (\Diamond_A A \wedge (A \rightarrow_A B))$ but the first of these is

$$(\dagger) \quad (A \wedge B) \vee (\Diamond_A A \wedge (A \rightarrow_A (B\|A)))$$

13.7 Modus Ponens and embedded conditionals

As I shall discuss in 13.7.2 there is good reason to believe that embedded conditionals express different propositions from conditionals that occur as the main connective of a sentence. It is helpful to give a systematic construction for what proposition a conditional expresses when it appears in different parts of a sentence. I shall now present such a construction.

13.7.1 My proposal for embedded conditionals

Let us define the *background* $B(u)$ of an utterance u to be either 1 or 0.

If u is not an embedded utterance, e.g. u is not uttered as part of $\ulcorner u$ and $v \urcorner$,[28] then $B(u) = \langle 0, \emptyset \rangle$.

- If u is a largest embedded sentence in some more complex sentence v, then $B(u)$ is identical to $B(v)$, unless v is a conditional:
- If u is embedded in $\ulcorner \text{if}\, u\, \text{then}\, v \urcorner$, and $B(\ulcorner \text{if}\, u\, \text{then}\, v \urcorner)$ then $B(u) = 0$.
- If u is embedded in $u' = \ulcorner \text{if}\, v\, \text{then}\, u \urcorner$, then $B(u) = 1$.

I define the background merely as a bookkeeping device, it is of no semantic significance. We can now give conditions on how to analyse conditionals appearing in various parts of sentences:

- Entailment is always analysed by $\rightarrow$, the strongest conditional.
- If u is of the form $\ulcorner \text{if}\, A\, \text{then}\, B \urcorner$ and $B(u) = 0$ then u should be analysed by $(A \wedge B) \vee A \rightarrow_A B$.[29]
- If u is of the form $\ulcorner \text{if}\, A\, \text{then}\, B \urcorner$ and $B(u) = 1$ then u should be analysed by $A \supset B$.

I have shown above (page 212) that if $\Diamond_A(A \wedge B)$ then $A \rightarrow_A (B \supset C) \equiv A \wedge B \rightarrow_{A \wedge B} C$. And so if A does not rule out B then $\ulcorner \text{if}\, A\, \text{then if}\, B\, \text{then}\, C \urcorner$ has the truth value as $\ulcorner \text{if}\, A \wedge B\, \text{then}\, C \urcorner$, which is as it should be.[30]

So for example 'if I toss a coin ten times it will land heads once' is false, but 'if I toss a coin ten times it will probably land heads once' is true. We should stipulate, as suggested above, that $\Diamond_\top A$ is true iff $p(A) \neq 0$. Now if we do not know whether $A \wedge B$ are true then we judge that $\ulcorner$if A then probably $B \urcorner$ exactly when we judge that † to be true which, assuming we do not know whether $A \wedge B$, we do if we judge that $p(B|A)$ is high.

[28] I use Quine's square quotes to allow for reference and quantification into a quotation.

[29] Or $(A \wedge B) \vee A \rightarrow_A B$, or $(A \wedge B) \vee (\Diamond_A A \wedge (A \rightarrow_A B))$ etc.

[30] Pollock ([**?**, p43]) rejects this, I remain unconvinced by his example.

There are cases where embedded conditionals are not material, for example if the grammar changes from indicative to counterfactual:

> If A then if it had been that B then C
> If A then if it were [then] the case that B it would be that C

or perhaps something more complicate like:

> If A then the situation would be such that if it were the case that B then C.

however, in general we can be safe to assume that embedded conditionals are material. Certainly the conditional 'if. . . as well, then. . . ' is material, as in

> If I go into the building then if I wear a hard hat as well, I will be safe.

13.7.2 Modus Ponens

It follows from the rules I have given that $A \rightarrow_A B$ and A entail B, i.e. it is a consequence of my theory that $\ulcorner \text{if } A \text{ then } B \urcorner$ and A entail B. This is an uncontroversial result, modus ponens is not commonly disputed. I wish now to discuss an argument that modus ponens is not, in general valid. If the argument is correct then the logic I have given is not correct (as an analysis of the logic of the conditional we actually follow). The argument shows that conditionals behave differently when embedded within other conditionals, I shall show how my theory can account for this.

The problematic inference is given by McGee [?] which he takes to be a counterexample to modus ponens. Opinion polls taken just before the 1980 election showed the Republican Ronald Reagan decisively ahead of the Democrat Jimmy Carter, with the other Republican in the race, John Anderson, a distant third. Those apprised of the poll results believed, with good reason:

- If a Republican wins the election, then if Reagan does not win Anderson will win.
- A Republican will win the election (i.e. Reagan).
- if Reagan does not win Carter will win (and not Anderson).

We are already able to give an explanation of what is going on:

According to the construction I gave for embedded conditionals in section 13.7.1, ⌜if A then if B then C⌝, is usually

$$A \rightarrow_A (B \supset C)$$

It is not complicated to see how this applies to McGee's apparent counterexample to modus ponens. Let A be 'a Repulican wins' and B be 'Reagan does not win'.

- a Republican wins the election $\rightarrow_A$ Reagan does not win $\supset$ Anderson will win.
- A Republican will win the election (i.e. Regan).
- Reagan does not win $\rightarrow_B$ Carter will win (and not Anderson).

And the final conditional need not be true and is not embedded in the initial premise. We therefore do not have a counterexample to Modus Ponens.

13.8 Concluding remarks

I take myself to have shown here that a proof theoretic account of analyticity can be applied even to an topic as problematic as the natural language conditional. The account sheds light on which truths of the logic of the conditional are analytic and which are not. I hope to have shown that there is a coherent account of the truth conditions of the conditional that treats the natural language conditional as a similar sort of entity to the strict conditional and the material conditional (which themselves are used in natural language).

With this I conclude that a proof theoretic account of analyticity has much to say about some very interesting words and constructions of natural language. I conclude that my theory of analyticity can account for the analyticity of some truths involving connectives, like 'if... then... ', that seem not to have an easy proof theory.

Chapter 14

The Thesis

I shall now postulate an account of how a definition is made. This content of this chapter is not definitive, I present it as a possible avenue for further research.

I think to obtain a full truth-by-definition account of analyticity we must provide an answer to these questions:

> Why should something that follows from definitions alone in this way be true? More generally, why should the proposition expressed by s follow validly from the propositions expressed by S just because s is analytic for S?

Ultimately what we want is a theory that answers the question of how we acquire a logical language for which a certain logical consequence relation is known to be valid.

To give an idea of what my proposed answer is, here is an example: the rule modus ponens has this form

$$\frac{\text{if}\, A\, \text{then}\, B \quad A}{B}$$

I claim that we do something that fixes the meaning of 'if. . . then. . . ' so that $\ulcorner \text{if}\, A\, \text{then}\, B \urcorner$ expresses a relation between $\|A\|$ and $\|B\|$, or more precisely, $\ulcorner \text{if}\, A\, \text{then}\, B \urcorner$ expresses a proposition that is true just in case a certain relation holds between $\|A\|$ and $\|B\|$. The relation is that of of $\|B\|$ following locally-validly from $\|A\|$ (local-validity is like validity, I define it in section 14.1). Now suppose that the premises of the modus ponens are true (i.e. express true propositions), then there really is a relation of local validity between $\|A\|$ and $\|B\|$. Furthermore, one of the premises (which we have assumed as true) expresses $\|A\|$, so $\|B\|$ is also true, and the conclusion of the modus ponens expresses $\|B\|$. Therefore, the conclusion of a modus ponens expresses a true proposition when its premises express true propositions.

On my view, the property of analyticity is therefore relative to a language. This adds the extra task of explaining why many propositions are universally expressed by analytic sentences across a wide variety of different languages, in particular it must be explained why logic is universal and is universally analytic.

14.1 Validity and local-validity

I assume that we possess innately (or acquire very quickly) concepts of validity and truth. I assume also that we possess a concept of local-validity. An inference from A to B is *locally-valid* when it depends on some other true assumptions Γ and the inference from Γ, A to B is valid. A special case of local-validity is truth (or perhaps we should say that local-validity is a generalisation on truth), for if A is true then an inference from nothing to A is locally-valid. Also, if an inference is analytically locally-valid, then it is *valid*.[1]

Truth an local-validity are the concepts we require to define the truth functional connectives, and truth and validity are the concepts we require to define strict implication (and related connectives).

Truth and validity are properties of propositions. Validity (and also local-validity) is a relation between the propositions expressed by the premises of an inference and the propositions expressed by its conclusion. Just as propositions satisfy or fail the property of truth in different possible circumstances, so do collections of propositions satisfy or fail the relation of local-validity in different circumstances.[2]

---

[1] The converse is not true, for example the inference from nothing to the proposition that Hesperus=Phosphorus is valid but not analytic.

[2] Whether the relation of validity also has this property is a tricky question. As I shall describe in chapter 10 we can formulate the logic of the strict conditional (meaning that the consequent follows validly from the antecedent) as $\Box(A \supset B)$ where $\Box$ is either the $S4$ or the $S5$ modality. I do not discuss which is more faithful to our innate concept of validity, thought I assume, tentatively, that it is $S4$ (in which case every valid argument is valid in all possible worlds, but some possible worlds have more valid arguments than this one).

14.2 A closer look at implicit definitions

I shall now suggest a more detailed story of how implicit definitions may be thought of as evolving out of explicit definitions rather than simply being alternative ways of defining terms.

There are two reasons why we cannot use explicit definitions to explain how we acquire a logical language and why some sentences/inferences of that language are analytically true/valid. Firstly, an explicit definition contains the term 'iff' or '=', and those are among the terms for which I seek to give an account. Thus an appeal to explicit definitions (where definiendum and definiens are linked by a logical connective) would be of little help. Secondly, in order to define the definiendum by means of an explicit definition we must already be able to express something with the same meaning: the definiens. So an explanation of our acquisition of a logical language in terms of explicit definitions would presuppose that we already acquired a logical language. The circularity is certainly vicious.

A natural stragey for resolving the first difficulty is to try to argue that the form of an explicit definition need not be entirely verbal. The point of an explicit definition is an association between two terms: definiendum and definiens. That is, an explicit definition merely to ensures that

> $\ulcorner$definiendum$\urcorner$ expresses the definiens

but this could be done without using the word 'iff'. I think it is enough that the speaker merely use the definiendum, on a particular occasion, with the intention of it meaning the same as when he uses the definiens. If he does such a thing then, I suggest, on that occasion he has fixed the meaning of the definiendum.[3]

The second difficulty can be resolved in a similar way. What I have in mind is that the definiens may be something that a speaker can think but not express. For example, a speaker may be able to think both A and B, he may realise this and then introduce a term 'and' (from the syntactic category of propositional connectives) so that

> $\ulcorner A$ and $B\urcorner$ expresses the local-validity/truth of: A, B.[4]

or perhaps a speaker may be able to make and discern reasoning from A to B and introduce 'if. . . then. . . '

[3] Although the speaker may have a hard time telling other people what $\ulcorner$definiendum$\urcorner$ means without verbally formulating the explicit definition.

[4] That is, $\ulcorner A$ and $B\urcorner$ expresses that $\|A\|$ (the proposition expressed by A) and by $\|B\|$ satisfy the property of truth.

$\ulcorner$if A then $B\urcorner$ expresses the local-validity of: $\begin{array}{c}\cancel{A}\\ \vdots\\ B\end{array}$.[5]

On the left hand side of these examples is a term ('and', 'if…then…') and on the right hand side is, not an expression, but an inferential structure. If a speaker can identify the inferential pattern in his own thoughts then he can define a term for it by saying, for example ' "and" expresses that' where by 'that' he refers to the inferential pattern. The situation is analogous to the case were we may define a proper name by identifying an object and saying ' "Julius" refers to that'. Of course, as I have just formulated it the first difficulty above arises again, it can be removed by treating the definition as made not by a verbal act '$\ulcorner$x$\urcorner$ expresses that' but by an intention to use $\ulcorner$x$\urcorner$ to express/denote that thing the speaker has identified.[6]

In overcoming this second difficulty the definitions of new propositional connectives have become of this form

(†) $\ulcorner Con(A_1 \ldots A_n)\urcorner$ expresses the local-validity of: $\Phi(A_1 \ldots A_n)$

where $\Phi(A_1 \ldots A_n)$ is an inference pattern (since inferences are made up of thoughts it is also a thought pattern). Since $\Phi(A_1 \ldots A_n)$ is a pattern rather than an expression it is not really a definiens, and so † is an implicit rather than an explicit definition.

But we can still apply (i), (ii) and (iii) from section 3.2.3 on page 48 to †. So (i) *Con* cannot occur in Φ, (ii) *Con* must be a single term (in † it has been specified as a propositional connective), and (iii) *Con* must be totally new to the language.

In the case of the logical constants, since we are explaining how they get their meaning when they are first introduced (iii) is met by assumption. We must now find definitions along the lines of † for the logical constants. The examples above for implication and conjunction seem initially plausible. Disjunction is problematic, a possibility is that disjunction is defined to express that case analysis is locally-valid, where we have an innate understanding of what case analysis is. Case analysis involves considering a number of cases and concluding anything that follows from all of them. Thus the definition for disjunction is

[5] That is, $\ulcorner$if A then $B\urcorner$ expresses that $\|A\|$ and $\|B\|$ satisfy the relation of local validity. A is crossed out to signify that it may be hypothesis rather than a belief (i.e. the speaker need not believe A in order to reason from it to B).

[6] I have a theory of how this works. I suggest that a theory in the spirit of Grice's intentional theory of meaning (from [**?**] and [**?**]) may be restated, not as a theory of what a word means, but as a theory of how the meaning of a word is fixed. I use such a theory to account for how an intention can fix the meaning of a term.

$$\ulcorner A \text{ or } B\urcorner \text{ expresses the local-validity of: } \frac{\begin{matrix}\cancel{A} & \cancel{B}\\ \vdots & \vdots \\ C & C\end{matrix}}{C} \text{ (for any } C).$$[7]

Another possibility for disjunction is that it is defined in terms of negation

$$\ulcorner A \text{ or } B\urcorner \text{ expresses the local-validity of: } \begin{matrix}\cancel{\sim A} \quad \cancel{\sim B}\\ \vdots \\ \bot\end{matrix}\ .$$[8]

where $\bot$ is absurdity.

14.2.1 Inference rules

Looking closely at the example implicit definitions above we can see that they can be reformulated in terms of the familiar inference rules. Reading

$$\ulcorner \text{if } A \text{ then } B\urcorner \text{ expresses the local validity of: } \begin{matrix}\cancel{A}\\ \vdots \\ B\end{matrix}$$

from right to left we get this rule:

$$\frac{\begin{matrix}\cancel{A}\\ \vdots \\ B\end{matrix}}{\text{if } A \text{ then } B}\ I$$

and reading it from left to right we get:

$$\begin{matrix}\dfrac{\text{if } A \text{ then } B}{A}\ E\\ \vdots \\ B\end{matrix}$$

a more familiar way of putting the rule E is:

$$\frac{\text{if } A \text{ then } B \quad A}{B}$$

[7]That is, $\ulcorner A$ or $B\urcorner$ expresses that $\|C\|$ is true for any C such that both $\|A\|$ and $\|B\|$ satisfy the relation of local-validity with $\|C\|$.

[8]That is, $\ulcorner A$ or $B\urcorner$ expresses that $\|{\sim}A\|$ and $\|{\sim}B\|$ together satisfy the relation of local-validity with $\|\bot\|$.

Notice that the rule E allows us to infer no more from $\ulcorner \text{if } A \text{ then } B \urcorner$ than is required to infer $\ulcorner \text{if } A \text{ then } B \urcorner$ using the rule I. In short, these rules are in harmony (in the first sense), in general they will be in harmony in the second sense as well.

Now, in addition to its relation to conservativeness, the significance of harmony (in its second sense) is this. If Con is a connective for which there are harmonious rules then we can express whatever $Con(A_1 \ldots A_n)$ expresses without using Con.[9]

In other words, a harmoniously defined connective expresses nothing over and above a certain inferential pattern. To go further, I conjecture that a connective that expresses nothing over and above an inferential pattern will have harmonious inference rules. That is, any connective that is defined by legitimate implicit definitions of this form:

(†) $Con(A_1 \ldots A_n)$ expresses the local-validity/truth of: $\Phi(A_1 \ldots A_n)$

will have harmonious inference rules. Furthermore, I suggest that any connective the inference rules for which are harmonious may be defined by definitions of the form of †.

Thus, if we want to check whether what is apparently a logical connective really is a *logical* connective (a connective definable entirely by its inference rules) we must check that it expresses nothing over an above some structural property of inferences. That is, we must check that the connective can be defined in the form of †. To check this we must show that the connective (or system of connectives) satisfies a normal form theorem when put together with all the other logical connectives (to show that each logical connective is conservative over all the others, so we may conclude that each connective is legitimately defined in the logical system as a whole).[10]

14.2.2 My account does not involve inferential roles

My account is not that the logical connectives are defined by their inference rules. The meanings of the logical connectives are fixed by some action the speaker performs, a baptism of some inferential structure. The meanings of logical connectives are not fixed by their inferential roles. My account is that the connectives are defined not by any disposition to use certain rules

[9]For example, instead of asserting $A \wedge B$ we can assert A and assert B.

[10]This should not be an unacceptable form of holism, for it does not entail that the meaning of any one connective is dependent on the others. It entails only that the *definition* of a logical connective can regarded as a legitimate definition only in the context of the definitions of the other logical connectives.

but by a few initial definitional uses of the rules (or perhaps even a single use).

The inferential role theory has too many difficulties. The challenges of Kripke and Quine, that behaviour underdetermines what rules we are to regard as the genuine definitional rules, are I think insuperable. The only response to them (on the implicit definition account) is to regard the implicit definitions as made explicitly. That is, from a early age a speaker must state some rules for the logical connectives and assert explicitly something like 'this connective is hereby defined by these rules'. Put this way is safe to assume that this does not happen. I try to formulate a more plausible account of how an implicit definition may be made explicitly.

Since, on my theory, the inference rules are a consequence of the explicitly made implicit definitions of the logical constants we can account for error and underdeterminacy. A speaker is in error when the rules he uses are not in accordance with the definition he has made of the connective in question. Similarly it is no matter that our behaviour underdetermines what rules we are using for our connectives for the rules we *should* be using are determined entirely by the initial definitions of the logical connectives.

14.2.3 My account is not metalinguistic

Neither is my theory that $\ulcorner \text{if } A \text{ then } B \urcorner$ expresses a meta-linguistic proposition like: an inference with $\ulcorner A \urcorner$ as the premise and $\ulcorner B \urcorner$ as the conclusion is locally-valid. My theory is that $\ulcorner \text{if } A \text{ then } B \urcorner$ expresses a relation between the propositions expressed by A and B, that relation being local-validity.[11]

14.3 My account

My account is that we define the logical connectives by means of implicit definitions that we make explicitly. That is, we do not express a definiens by which the definiendum is defined, but, by means of an explicit action, make an implicit definition of the logical connectives.

For example, I claim we possess an ability to think conjunctively. If we use the word 'and', just once, with only the intention of causing others to think conjunctively, then we fix the meaning of 'and' to have the truth conditions of conjunctive thoughts.

But what does it mean to 'think conjunctively' and what is a 'conjunctive thought'? It may seem that my account answers the charge of circularity

[11] I of course deny that validity, truth and local-validity are meta-linguistic concepts.

at the cost of allowing a charge of unhelpfulness. Is it not that 'thinking conjunctively' is in equal want of explanation as the term 'and' and its meaning? For sure it does, but the explanation is now quite simple.

In order to think conjunctively we need only think two thoughts. In order to think inferentially we need only draw an inference. In order to think disjunctively we need only perform a case analysis.[12] Now, anything that can think at all, I claim, can think more than one thing. Moreover, anything that can think at all can draw inferences. Furthermore, anything that can perform higher level thoughts (e.g. humans) can have some thoughts that are beliefs and others that are suppositions.

I do not claim that the ability to make valid or rational inferences is a necessary condition of thought, all I claim is that something does not think unless it can have a number of thoughts and beliefs and suppositions (true or false) and make inferences (valid or invalid, rational or irrational).

In short, I claim that we possess innately psychological analogues of the symbols of the natural deduction system. Put another way, I claim that the Prawitz natural deduction system provides a good formalisation of some minimal capacities a creature must have to be a thinker.

I assume that we possess innately (or at least develop very quickly):

- the ability to make, and recall as such, a chain of inferences i.e. a chain of reasoning (not necessarily with any degree of rationality)
- the ability to apply, and recall as such, reasoning by case analysis
- the ability to call to mind one belief out of many and recognise collections of beliefs.
- the ability to recognise suppositions and distinguish them from beliefs.
- enough of a concept of truth to know that some propositions are true and to know that some inferences preserve truth.[13]

It is a matter of empirical testing to see if these are indeed innate capacities and how they manifest themselves as innate (e.g. it may be a matter of hardwiring). The abilities above seem to be ones that we must assume we

[12] A case analysis for A and B involves concluding as true anything that we can derive from *each of* A and B.

[13] This does not entail that we have enough knowledge of truth to be able to say exactly what it is. Neither does it entail that truth is some innate irreducible primitive. The debate on the nature of truth is a side issue here. I am not concerned with what truth is, I am more interested in why some sentences are analytically true.

have *at the very least*, it is the strongest of all sceptics (and a self-refuting one at that) who doubts that we have these capacities. I claim they are capacities that we cannot alter without impairing our capacity to think and are innate.[14]

A good formal analysis of these capacities is given by the structure of the Prawitz natural deduction system. At least, the structure of the more natural Prawitz deduction system I sketch in 1.2.

To understand the natural deduction rules we must understand that it is a deduction system, this is the analogue of our ability to expect that something is preserved. We must understand the horizontal line,

$$\frac{\Gamma}{\Delta}$$

this is the analogue of our ability to order beliefs as following from each other. We must also understand the vertical dots,

$$\begin{array}{c}\Gamma\\ \vdots\\ \Delta\end{array}$$

this is the analogue of our ability to recall chains of reasoning as chains of reasoning. We must be able to understand what Δ and Γ are and contain, this is the analogue of calling to mind one belief from many (recognised as many). We must also understand what crossing out formulae and the bookkeeping of superscripts is (i.e. discharging assumptions), this is the analogue of distinguishing suppositions from beliefs. We must also be able to understand that deductions may function as minor premises (e.g. in the rule $\vee E$), this is the analogue of case analysis.

So for example, we possess psychological analogues for, among other things, the horizontal line and the vertical dots:

$$\_\_ \quad \vdots$$

But we, as higher level thinkers, do more than possess these psychological properties, we can discern them. That is we can recognise (at least some of) our beliefs and suppositions as such, and we can recognise our inferences as such. Whereas, say, a cat might merely have a sequence of thoughts that

[14] In a way, the framework constituted by these capacities is revisable in that we can remove some of them (a heavy blow to the back of the head usually does the trick). However any person without all of these capacities is defective and is excluded as an example in my discussion.

constitute an inference, I, a higher level thinker, can have such a sequence of thoughts and recognise it as an inference. Reflection alone, probably the first piece of introspection we ever do, allows us to recognise that we have these psychological properties (of possessing many beliefs, drawing inferences etc.).

Furthermore I claim that we possess innately a concept of truth and concepts of truth preservation (i.e. validity and local-validity), for otherwise we could not distinguish beliefs from suppositions or inferences from other sequences of thoughts. For example a belief is a supposition we hold as true, and an inference is a sequence of thoughts where we hold that each successive thought preserves the truth of its predecessor.[15]

So we can put all of this together. The word 'and' obtains its meaning by being used to express the truth conditions of conjunctive thinking. We identify conjunctive thinking by introspecting many of our beliefs. We then use the word 'and' with the intention that (regardless of what A and B are) $\ulcorner A$ and $B\urcorner$ produce a response in an interlocutor of possessing these (many) beliefs: A, B.

More generally, a logical constant is defined by an definition of this form:

$$Con(A_1 \ldots A_n) \text{ expresses: } \Phi(A_1 \ldots A_n)$$

where the left hand side is a sentence involving a term C of the syntactic category 'sentential operator', and on the right hand side is some complex thought pattern relating $A_1 \ldots A_n$.[16] Examples are:

$$\ulcorner A \text{ and } B\urcorner \text{ expresses (the local-validity of): } A, B$$

and

$$\ulcorner \text{if } A \text{ then } B\urcorner \text{ expresses the local-validity of: } \begin{array}{c} \cancel{A} \\ \vdots \\ B \end{array}$$

These are implicit definitions as the 'definiens' is a pattern rather than an expression.[17] But the definitions are nonetheless explicit as there is an explicit intention of using a word in a certain way.

---

[15] As considerably more mature thinkers we possess many, more subtle, concepts of truth and inference. The innate concepts of truth and inference are quite basic, and need be enough only to give us the ability to discern beliefs from suppositions and inferences from unconnected thoughts.

[16] I take it that we possess an innate (universal) grammar that provides us knowledge of the various syntactic categories). I also take it that we possess innately a knowledge that words can be used to express things and have, innately, a concept of what it is to express a concept or some truth conditions with a word.

[17] We identify it by a piece of introspection, e.g. introspecting the presence of two beliefs

14.3.1 Classical logic

The famous inference rules for classical logic do not normalise. The problem is double negation elimination.

$$[DNE]\frac{\sim\sim A}{A}$$

Consider a negation that is introduced and then eliminated by DNE

$$\cfrac{\cfrac{\begin{array}{c}\cancel{\sim A}^{1}\\ \vdots\\ \bot\end{array}}{\sim\sim A}\;\sim I(1)}{A}\;DNE$$

it seems that there is no general way of obtaining a direct deduction of A out of a deduction from $\sim A$ to $\bot$:

$$\begin{array}{c}\cancel{\sim A}\\ \vdots\\ \bot\end{array}$$

Effectively, we must show that this inference pattern:

$$A$$

is locally valid just in case this one is:

$$\cfrac{\begin{array}{c}\cancel{A}\\ \vdots\\ \bot\end{array}}{\bot}$$

There are at least two ways to achieve this

1. Devise a specific system which is equivalent to the usual system (of inference rules) for classical logic which just so happens to be such that there is a direct deduction of A whenever there is a deduction of $\bot$ from the negation of A. An example of this is the logic for Sheffer Stroke.

A, B for the definition of conjunction. I am inclined to hold that we can refer to such thought patterns in much the same way that we can pick out any structural property of real objects. For example we are able to notice a structural property, e.g. a linear order, and say '*that* is how people are ordered in a queue'.

2. Enhance the structure of the natural deduction so that, in general, there is a direct deduction of A whenever there is a deduction of $\perp$ from the negation of A. An example of this is the Restart rule.

Ultimately, I favour the generality of the second strategy and adopt the restart rule.

14.4 Knowledge of inference and analyticity

I now distinguish between what I call a *logic user* and a *logic knower*. A logic user knows logical truths and preserves knowledge through logical inference (that is, ceteris paribus, if a logic user knows A and infers B from this deductively, then he knows B).[18] A logic knower is a logic user who knows that logic is analytic. I shall now explain how each is possible.

I assume that analysis of knowledge as at least justified reliable belief will do for the sorts of belief I discuss here.

14.4.1 Logic users

Since harmonious introduction and elimination rules for a connective match a structure a speaker has already identified, he will naturally use those rules as the basic rules for that connective. Furthermore, I claim that any belief the speaker has in the local-validity of those rules is reliable, for such a belief is a direct consequence of that connective expressing (the truth conditions of) a structural property of inferences, which then guarantees that those rules are valid (as well as that they are locally-valid). Finally, the belief that the inference rules are locally-valid is justified, the justification may come either from the speaker's interaction with other users of the same logical connective and from the speaker's association of that logical connective with a certain inferential structure.[19]

This is my explanation of how a logic user is possible and is to be expected.

14.4.2 Logic knowers

Note that a logic user, on any occasion, knows that a logical truth is true or that a logical inference is locally-valid. It would take a further piece of

[18] I take it that Nozick's famous rejection of the transmission of logical knowledge through known entailment is a problem for Nozick rather than for me.

[19] For example, if the speaker always associates $A \wedge B$ with these beliefs: A, B; then the speaker is justified in inferring A and B from $A \wedge B$.

introspection not only to notice that an inference is locally-valid by definition.

Although normal form theorems (i.e. normalisation theorems) can be hard to prove, it is often easy to see that they are provable. That is, it is easy to see that there is a symmetry between the introduction and that elimination rules and elementary deduction reductions can be made. We do not need a proof of a formal truth to know that it is true, Knowledge does not require infallible proof. Thus, to know that a connective derives is meaning entirely by definition it is enough introspection, I claim, to note what our logical rules are and see the symmetry between the introduction and elimination rules, for example we may see they are in harmony in its first sense (see pageref 3.3.1). From this a speaker obtains knowledge of normalisation and harmony and thence knowledge that the terms in question are legitimately defined. This is how a logic knower is possible.

In many respects we are logic knowers, but perhaps sometimes we cannot discern the rules we use or see so easily that the terms in question are defined by legitimate definitions. I claim that we are knowers of all the logics of this thesis for which I provide a normalisation theorem. Some logical connectives are not entirely logical, for example the natural language conditional (which is neither strict nor material) is partly defined in terms of (crudely put) the closeness relation on possible worlds which is something we discover rather than define. For such connectives, especially where the non-logical elements of them are hard for us to define, we may be logic users rather than knowers.

14.4.3 Universality

Also I claim to have given an explanation of the universality of logic. The same propositions expressed by the theorems of the logic of my language are analytic in so many different languages too because they all use the same logic. The logic is the same because it is defined out of innate abilities and capacities than any thinker must possess to be a thinker. Thus, if there are aliens, their logic should be the same as ours.

14.4.4 The response to Kripke and Quine

I sum up the challeneges of Quine and Kripke with this question:

> What matter of particular fact was there that gave certain words the particular meaning they do so that sentences involving them are analytic (or have a particular meaning at all)?

My response is this:

> Some of our terms are, at some point in our lives, used with a certain intention. This use of the word was an event the occurrence of which is a matter of particular fact. In many cases, that use gives the word a particular meaning. In the case of logical connective Con, $\ulcorner Con \urcorner$ is used with the intention of expressing the local validity of reasoning Con-ly.
>
> I argue that if the rules for Con are part of a harmonious logical system then anything that can reason at all can reason Con-ly and recognise such reasoning as Con-ly reasoning.[20]

14.4.5 A comment on semantics

I claim that truth is an innate concept, furthermore this concept is central to our ability to define logical connectives. Without the concept of truth we could not even discern inferences, let alone inferential patterns, by which we define the logical connectives. For this reason I suggest that the correct semantics for are logical connectives be a truth conditional semantics. An analysis of the metaphysics of truth would tell us how literally we should interpret the more familiar semantic analysis of our logical connectives. For example, I shall present a modal logic and give a possible world semantics for it. If truth is, say, satisfaction at a possible world (read at face value, like Lewis does) then we may understand this possible world semantics literally. However, if possible worlds are merely a convenient fiction then we must find some other interpretation of them (or a different semantics entirely). However, what must always remain is truth, any semantics we give must be truth conditional.

I shall not discuss further how the semantics should be interpreted aside, I am more concerned about what the truths of logic are than exactly what semantics should be given to them.

[20] So anything that can think at all can reason conjunctively, disjunctively, negatively and conditionally.

Appendix A

Second order logic and plurals

The conclusions of this appendix are these:

- Natural language makes use of higher order quantifiers which range over *all* classes of elements of an approprate domain. I discuss a higher order extension of classical logic which can handle sentences such has 'some critics admire only each other'.

- There is no requirement to intepret the higher order quantifiers of natural lanugage as quantifying over *objects* (like sets or other sorts of plural entity). We can handle sentences like 'some men raced each other', 'John and Mary wrote a book (together)' without an ontology of plural entities.

- We can handle sentences like 'some men raced each other', 'John and Mary wrote a book' in a first order logic (but not 'some critics admire only each other').

I obtain these conclusions by discussing some issues surrounding the expressive limitations of first order logic. A famous limitation of first order logic is its inability to handle quantification and predication over plurals. For example, we cannot use first order logic to quantify over collections (for example the Geach sentence is inexpressible in first order logic, see section A.2.1). Furthermore first order logic, apparently, cannot handle sentences like 'John and Mary lifted a piano' (where it appears we are predicating lifting a piano of a collection constituted by John and Mary).

To begin with I shall investigate here how we should extend first order logic to overcome these limitations. I shall present a higher order logic with quantifiers that may be interpreted as ranging over sets (though I argue that this is not how they should be interpreted).

I shall then discuss in more detail the Geach sentence ('some critics admire only each other') and its relation to the higher order logic obtained by adding an operator that expresses transitive closure. I will show that we can use the ancestral operator to express the (structure of) the Geach sentence and vice versa.

Finally, the higher order logics I will have discussed do not have any mechanism for plural reference. It seems that sentences like 'some critics lifted a piano (together)' involve reference to a plural entity, but such an analysis unavailable to me as I suggest we should not interpret the higher order quantifiers as ranging over plural entities. I suggest that we interpret such sentences as involving a *plural predication* over some things (rather than a single predication of a plural thing).

It turns out that if we work with plural predications (rather then plural reference) then we can handle sentences like 'some critics lifted a piano' in *first order logic.*

A.1 Higher order quantification

A.1.1 The rules

A good reason for adding machinery for higher order quantification is to improve expressive power. It is not simply that the limitative results in first order logic yield non-standard models of things which we try to express with it. There are sentences which our basic logic should be able to express that first order logic cannot even make an attempt at. An example of such a sentence is due to Geach:

Some critics admire only each other

Geach sentences are of interest in their own right and I shall discuss their properties in more detail in A.1.3.

The simplest way improving expressive power is to allow quantification over predications. We can add predicate variables to the language such that for each $n > 0$ there are n-ary predicate variables $\alpha, \beta, \gamma \ldots.$

Let $A[x/t_1 \ldots x/t_n]$ be a formula in which each t_i is free for x in A. Furthermore, let $(\ldots A \ldots)$ be a formula in which A occurs as a subformula (variables free in A may be bound in $(\ldots A \ldots)$). Finally let $A[\alpha/\beta]$ be result from A by replacing all occurrences of $\alpha(t_1 \ldots t_n)$, not in the scope of a quantifier $\exists\alpha$, by $\beta(t_1 \ldots t_n)$ (for any terms $t_1 \ldots t_n$).

$$\frac{(\ldots A[x_1/t_1, \ldots, x_n/t_n] \ldots) \quad A[x_1/t'_1, \ldots, x_n/t'_n]}{\exists\alpha(\ldots \alpha(t_1 \ldots t_n) \ldots)}\, I$$

Provided the t'_i are free for x in A

$$\frac{\exists\alpha A \qquad \begin{array}{c} \cancel{A[\alpha/\beta]}, \cancel{\beta(c_1 \ldots c_n)} \\ \vdots \\ C \end{array}}{C}\, E$$

where the c_i are constants and neither β nor the c_i occur in any formulae or assumptive rule applications on which C depends, except $A[\alpha/\beta]$ and $\beta(c_1 \ldots c_n)$, nor in $\exists\alpha A$ nor in C itself.

In these rules α and β are n-ary predicate variables ($n > 0$). Further the extra premise that of $A[x_1/t'_1, \ldots, x_n/t'_n]$ be deducible in the introduction rule ensures that there really is something that is A when we abstract it away to α. That is, $\exists\alpha A$ says that there is a *nonempty* collection of things satisfying A.[1]

I do not give rules for the universal quantifier as, in general, when we use 'all' we are using the first order quantifier. It is not necessary and is far more trouble than it is worth. If a universal quantifier is required we may define it in terms of the existential quantifier $\sim\exists\alpha\sim\ldots$.

We can now formalise a sentence like

Some critics are evil

as

$$(\dagger)\ \exists\alpha\forall x[\alpha(x) \supset (Cx \wedge Ex)]$$

[1]We can supplement these rules by adding additional predicates for phrases like 'John and Mary', 'Russell and Whitehead' etc. we add a unary predicate $\{t_1 \ldots t_n\}$ for every such string of terms

$$\frac{}{\{t_1, \ldots, t_n\}(t_i)} \text{ where } 1 \leq i \leq n. \qquad \frac{\{t_1 \ldots t_n\}(t) \qquad \begin{array}{c} \cancel{t = t_1} \\ \vdots \\ C \end{array} \quad \ldots \quad \begin{array}{c} \cancel{t = t_n} \\ \vdots \\ C \end{array}}{C}$$

The details of this get tricky when we allow quantification into the predicate $\{t_1, \ldots, t_n\}$. For these reasons I avoid developing a formal theory of these predicates, also, my account of plural predication does not require them.

where E is 'dots is evil'. This deduction

$$\dfrac{\exists\alpha\forall x(\alpha(x)\supset(Cx\wedge Ex)) \qquad \dfrac{\dfrac{\cancel{\beta(c)}^1 \qquad \dfrac{\cancel{\forall x[\beta(x)\supset(Cx\wedge Ex)]}^1}{\beta(c)\supset(Cc\wedge Ec)}\,\forall E}{Cc\wedge Ec}\,\supset E}{\exists x(Cx\wedge Ex)}}{\exists x(Cx\wedge Ex)}\,E(1)$$

shows that † entails that some critic is evil. Also $\exists x(Cx\wedge Ex)$ entails †

$$\dfrac{\exists x(Cx\wedge Ex) \qquad \dfrac{\dfrac{\dfrac{\dfrac{\cancel{Cc\wedge Ec}^1 \quad \cancel{x=c}^2}{Cx\wedge Ex}\,{=}E}{x=c\supset(Cx\wedge Ex)}\,\supset I(2)}{\forall x(x=c\supset(Cx\wedge Ex))}\,\forall I}{\exists\alpha\forall x(\alpha(x)\supset(Cx\wedge Ex))}\,I}{\exists\alpha\forall x(\alpha(x)\supset(Cx\wedge Ex))}\,\exists E(1)$$

in this deduction we abstracted away $x=c$ by $\alpha(x)$ in the application of the higher order introduction rule.

A.1.2 Normalisation

Normalisation follows in the same way as with the first order existential quantifier. Suppose $\exists\alpha$ is introduced and then eliminated

$$\dfrac{\dfrac{\begin{matrix}\vdots\\(\dots A[x/t_1,\dots,x/t_n]\dots)\end{matrix} \qquad \begin{matrix}\vdots\\A[x_1/t'_1,\dots,x_n/t'_n]\end{matrix}}{\exists\alpha(\dots\alpha(t_1\dots t_n)\dots)} \qquad \begin{matrix}\cancel{(\dots\beta(t_1\dots t_n)\dots)} \quad \cancel{\beta(c_1\dots c_n)}\\ \vdots\\ C\end{matrix}}{C}$$

β functions as any other (atomic) predicate as far as the inference of C from $(\dots\beta(t_1\dots t_n)\dots)$ is concerned. Our choice of β allows us to substitute any occurrence of $\beta(d_1\dots d_n)$ by $A[x/d_1,\dots,x/d_n]$ for any terms $d_1\dots d_n$. Our choice of constants c_i allow us to substitute them for any terms we like, in particular the t'_i. Thus we obtain a direct inference of C from $(\dots A[x/t_1,\dots,x/t_n]\dots)$ and $A[x/t'_1,\dots,x/t'_n]$:

$$\begin{matrix}\vdots & & \vdots\\ (\dots A[x/t_1,\dots,x/t_n]\dots) & & A[x/t'_1,\dots,x/t'_n]\\ \vdots & & \\ C & & \end{matrix}$$

A.1.3 Plural quantification

We can use the extra expressive power to formulate the Geach sentence:

$$\exists\alpha\forall y((\alpha(y) \supset (Cy \wedge \forall z(Ayz \supset \alpha(z)))$$

or perhaps

$$\exists\alpha\forall y((\alpha(y) \supset (Cy \wedge \forall z(Ayz \equiv \alpha(z)))$$

if we understand the Geach sentence as 'some critics admire all and only each other'.

The usual interpretation of $\exists\alpha A$ is that there is a collection, or class, of elements that satisfy A. $\alpha(t)$ satisfied when the element assigned to t is a member of the class assigned to α. We can also choose to interpret α as ranging over the syntax, so that $\exists\alpha A$ receives a substitutional interpretation along the lines of 'there is a description (or predicate) which when substituted for α in A yields a true sentence'. The substitution interpretation of higher order quantification is far too tied to the syntax, we might assert the Geach sentence without having anything to substitute for α to form a true sentence not involving the higher order quantification. It seems that if we are to be able to use the rules above as a definition of higher order quantification we must have developed (or possess innately) something of a concept of class or plurality. There is an alternative in the literature to understanding such quantifications in terms of sets, we may take as a primitive, plural quantification (e.g. Boolos [**?**], Hossack [**?**]). Our cognitive capacity that allows us to define the higher order quantifiers need not be, on this view, a concept of a set, rather a concept of plural quantification. With plural quantifiers as the primitive we may then develop much of set theory (without taking a concept of set as the primitive).

Notice that the higher order rules I present do not give us a language of set theory. In particular we cannot reason with predicates like we reason about objects, for example we cannot predicate α (as in $F\alpha$) like we can predicate a name a (as in Fa). For this reason we do not need full set theory in order to understand the rules of higher order quantification that I present. Standard set theories are theories of pluralities (a.k.a. classes) where these pluralities are themselves treated as objects of membership to other pluralities. In order to understand set theory we must develop a more advanced concept of a plurality as an object, this is far easier said than done especially when we try to keep the concept consistent (as Frege discovered).

When we do not think of pluralities as objects of plurality themselves we do not have a set theory, we have an elementary theory of plurality. If we allow quantification over pluralities then we have an elementary theory of plural quantification. As far as I can see, if the higher order quantification remains quantification over classes of objects, and we do not bring each class down into the domain of quantification as an object (i.e. as long as we do not develop set theory), then there is no difference between taking classes as the primitive concept rather than plural quantification.

Thus I claim that we use the higher order rules above to define higher order quantifiers in accordance with a pre-existing concept of plurality. Of course, we can use these rules to develop a theory of sets. We can give axioms that state that for every plurality there is a corresponding object and thereby enrich the domain of quantification with considerably more entities (a full domain of sets). But such a development is a theory, a theory which has its uses in some but not all applications (e.g. the axiom of choice is useful to some applications and not to other applications). A full set theory is not required to provide an understanding of higher order quantification. Finally I wish to touch here on a debate about the interpretation of higher order logic. We may form a models, call them H-models, for our higher order logic the domains of which feature two sorts of object, urelements (non-sets) and pluralities (i.e. sets). As long as we have enough sets in the domain to be associated with every formula A of the language (that can occur as a premise of the higher order introduction rule) then the logic will be complete for these models.[2] We can then interpret higher order quantification as first order quantification over the domain of sets. Now all the limitative results of first order logic apply to the logic, for example if the syntax is denumerable then the H-models may be denumerable. H-models provide an unacceptable interpretation of higher order quantification in general as we can, if the language is denumerable, keep the H-models denumerable even though the number of pluralities should be far greater the number of elements. For example if there are denumerably many urelements there should be uncountably many pluralities, but according to the H-models there are not. We might worry that, since we can always interpret the rules above in terms of H-models we cannot to justice to higher order logic as we cannot characterise the 'real' models of higher order logic.[3] I shall now discuss why I do not find this worrying.

Firstly the existence of alternative interpretations of our terms does not alone undermine or raise any sceptical doubts about their meanings. We

[2] This is called the Henkin semantics for second order logic, see Shapiro [?].

[3] In section II of [?] Shapiro discusses a debate that might arise between someone who insists on interpreting higher order rules in terms of H-models and someone who intends a far richer (second order) interpretation of the higher order rules above.

do not have to characterise something exactly in order to refer to it or talk about it. I do not need to produce a complete theory of the arithmetic that characterises their structure up to isomorphisms in order to give appropriate meaning to the word 'one'; I do not need to develop a full set theory in order to know what a set or class or plurality is; I do not need to develop a full psychology of human interpersonal relationships for my word 'love' to mean (or refer to) love. Secondly the H-models work by tailoring themselves to a fixed language, that is we can provide a first order interpretation of the higher order introduction an elimination rules only by treating them as schematic over a fixed language. This is clearly not what the rules are, the rule is *intended* to be general over any (extension of the) language we speak.[4] The inference rules should be interpreted far more generally than any first order interpretation of them.[5] Thirdly the first order interpretations of higher order logic require pluralities to be thought of themselves as objects (in order for them to be in a domain of first order quantification). As I have suggested above such an understanding of pluralities is quite an advancement on the elementary concept of plurality that allows plural quantification. Thus the first order interpretation should be thought of as an advanced theoretic development of higher order logic rather than its fundamental interpretation.[6]

A.1.4 Plural predication

Not only do we quantify over pluralities we seem also to predicate over them. For example

Russell and Whitehead wrote Principia Mathematica

has a reading that is not equivalent

[4] In the same way that the rules for conjunction are not for using $\wedge$ to connect formulae formed from a fixed syntax, they are rules for connecting formulae of any syntax we choose to work with.

[5] Our intention to use the rules so generally is the fact that determines the incorrectness of the first order interpretation.

[6] An answer to the question of which interpretation is correct does not affect the proof theoretic properties of higher order logic. Even though plural quantification really does quantify over *all* classes, the higher order logic I present is very weak. In order to obtain a full set theory we must add extra axioms and rules, in short, we must develop a theory. Set theory *is not* the basis of our understanding of plurality, it is a theoretical advancement of a concept we already possess (a theoretical advancement we can choose to use or reject at will, there are many set theories).

> Russell wrote Principia Mathematica and Whitehead wrote Principia Mathematica

Similarly there is a reading of

> Some critics lifted a piano

the truth conditions of which are of a collective piano raising action rather than a number of individual actions.

It is possible to develop a theory of predication where pluralities can take an argument position of a predicate. Such a line would defeat all that I have argued previously about the concept of plurality being weaker than the concept of a set. I cannot interpret a predication Fa unless a is some kind of object (abstract or real), that is I cannot give a sensible interpretation of Fa unless a is something that a first order quantifiers can range over. If we allow plural predication (e.g. $F\alpha$) then we must be able to treat pluralities like objects. But to do so would overturn my arguments above about the interpretation of second order logic.

Unless my mind is opened to other interpretations of plural variables occupying argument positions of predicates, I must seek another way of characterising plural predication. The method I propose arises from my analysis of Geach sentences in the context of the limitations of first order logic. I now turn to this analysis, my account of plural predication is contained within it. The idea is that we should treat plural predication, not as a predication of a normal predicate over a plural entity, but as a predication with a plural predicate over a number of (unitary) entities.

A.2 First order enhancements

A.2.1 Expressive difficulties of first order logic

A reason for accepting a second order logic over a first order logic is the failures of expressive power of a first order language. Intuitively we are capable of expressing sentences like

(F) There are finitely many people

(A) John is my ancestor

(G) Some critics admire only each other (the Geach sentence[7])

[7] Amusingly this is occasionally simplified to

(C) John lifts some stones (collectively)

The usual language of first order logic cannot express any of these. The first three at least fall foul of the *compactness theorem*.[8] The compactness theorem states that

> if S is a collection of sentences and every *finite* subset of S is consistent (has a model) then so is S.

Since every proof is finite, if S is inconsistent then the proof of this is finite. Take all the members of S required for this proof and we have an inconsistent finite subset of S. So if there is no such subset then S is consistent.

Now it is easy to express for each number n a first order sentence meaning 'there are at least n P', where P is any predicate. Consider the set S of sentences:

There are finitely many people

There is at least 1 person
There are at least 2 people
There are at least 3 people
$\vdots$

Every finite subset of this is consistent. But the whole set S is not consistent. Therefore (F) above is not expressible in a first order language.

Now consider the set S of sentences:

John is my ancestor
John is not a parent of mine
John is not a parent of a parent of mine
John is not a parent of a parent of a parent of mine
$\vdots$

Some things admire only each other

which has a very simple first order formulation. Since there cannot be anything outside the domain of 'thing' it is true exactly when something exists! As long as admiration is not transitive we will have a hard time with

Some things admire all and only each other

the problem is capturing the transitive closure of the non-transitive relation.

[8] I have shown how to extend the logic to a second order language, the compactness theorem does not hold for such a logic.

Except the first, each is easily rendered into first order logic. Every finite subset of S is consistent, but S is not. Therefore (A) above is not expressible in a first order logic.

If the Geach sentence (G) is expressible in a first order language then so is its negation. Consider the set S of sentences:

> It is not the case that some critics admire only each other
> John is a critic who admires only critics
> John is a critic who admires only critics who admire only critics
> John is a critic who admires only critics who admire only critics who admire only critics
> $\vdots$

As the sentences get longer we have to go further down the chain of admiration before we find a non critic. Since each sentence is finitely long we can be sure the chain can end. So every finite subset of S is consistent, but S is not. Therefore (G) is not expressible in a first order language.

I shall discuss (C) later.

A.2.2 Geach sentences

The Geach sentence (G) is of interest. Let us look more closely at the Geach construction

> Some P(s), R only each other

where P is a unary predicate and R is a binary relation. A similarity between the proof that (A) and (G) are not expressible in a first order language suggests a link between the Geach construction and the following ancestral constructions

> y is an R-ancestor of x
> y is in the transitive closure of R from x
> there are $x_1 \dots x_n$ such that xRx_1, $x_iR_xi + 1$ and x_nRy

If R is a binary relation then let R^* be the transitive closure of R. That is if xR^*y then y is an R-ancestor of x. Now consider the sentence

$$(\dagger) \quad \exists x(Cx \land \forall y(xR^*y \supset Cx))$$

where R is '... admires... ' and C is '... is a critic'. Then † means that there is a critic who has only critics in the transitive closure of the admiration relation from him onwards. That is he admires only critics and they admire only critics and they.... Take all the critics who are admiration ancestors of this critic, that is a collection of critics who admire only each other.

So, if we have the power to express transitive closures then we can formulate the Geach sentence and Geach constructions.

To go the other way we need to enhance the Geach construction, as it stands we cannot use it to refer to any particular individual. Here is a minor enhancement:

> there are some P(s) that R only each other and x but not y is among them.

Now we can express the sentence y is an R-ancestor of x. Take the negation of the enhanced Geach construction:

> it is not the case there are some things that R only each other and x but not y is among them.

The transitive closure of the R relation from x is the smallest set that contains x and the members of which R only each other. So if there is no set, closed under R, that contains x and not y, then y is an R-ancestor of x.

It is easy to verify that the enhanced Geach construction may be expressed using R^*. Thus being able to express the two Geach constructions is equivalent to to being able to describe the transitive closer of a binary relation.

A.2.3 Ancestral relations

The ability to express the transitive closure accords us with great power. For example we can once and for all define the natural numbers. Instead of using the first order induction axiom we let R be such that

$$xRy \equiv y = s(x)$$

where s is the successor function. We then add the rule that

$$\forall x(Nx \equiv (x = 0 \vee 0R^*x))$$

where N is to mean '... is a number'.

Actually, with sufficient cunning we do not even need this much, the original Geach sentence can characterise the numbers without modification. We require only that we have enough rules for arithmetic for guarantee that

0 has no predecessor
Every number is 0 or is a successor (has a predecessor)
Only numbers are successors of things

these are easily rendered in first order logic. We then need only add a clause that

(†) It is not the case that some successors are predecessors of only each other

To see that this is sufficient it is enough to show that this entails second order induction. Suppose that $F0$ and $\forall x(Fx \supset Fs(x))$ and now hypothesise that that $\sim Fa$. Let P be '...is a predecessor of...'. Then

$$(Pyx \wedge \sim Fx) \supset \sim Fy$$

now consider $X = z : \neg Fz$. Since $F0$ is follows that $0 \notin X$. Furthermore if $x \in X$ then $\sim Fx$ and so $\forall y Pyx \supset \sim Fy$ and so $y \in X$, thus

$$\forall x \forall y((x \in X \wedge Pyx) \supset y \in X)$$

So X is a set of objects that are predecessors of only each other. But this is contrary to †, therefore the hypothesis is not true so Fa is true. Since a was arbitrary we may conclude that

$$[F0 \wedge \forall x(Fx \supset Fs(x))] \supset \forall x Fx$$

and since F was arbitrary we may conclude its second order generalisation.

Once we have a predicate N that really does mean 'dots is a number' and has no non-standard interpretation we can use usual recursive techniques to define pretty much anything we like. For example if F is a unary predicate let us define a binary predicate G such that

$G(F, 0)$	$=_{df}$	there are no F
$G(F, 1)$	$=_{df}$	there are no F or there is exactly one F
$G(F, n)$	$=_{df}$	$C(F$,n-1) or there are exactly $n-1$ F

and the right hand sides are easily rendered into first order logic. Since we are sure that this definition covers all the numbers (and there are no non-standard numbers lurking around for which C might not do what it is supposed to do), then

$$\exists x(Nx \wedge C(F, x))$$

means 'there are finitely many F'.

A.2.4 Ancestral sentences

We can give inference rules for the ancestral operator. I give them here for the ancestral relation on a binary relation and I claim these are rules we do use:

$$\frac{tRt_2}{t_1R^*t_2}\,{*I}\qquad\qquad \frac{t_1R^*t_2 \quad t_2Rt_3}{t_1R^*t_3}\,{*P}$$

$$\frac{t_1R^*t_2 \quad A[x/t_1] \quad \begin{matrix}\cancel{A[x/c]}^n \quad \cancel{cRd}^n \\ \vdots \\ A[x/d]\end{matrix}}{A[x/t_2]}\,{*E(n)}$$

provided t_1 and t_2 are free for x in A and provided c and d do not occur in any formulae or assumptive rule applications on which $A[x/d]$ depends (except $A[x/c]$ and Rcd), nor in A

The elimination rule $*E$ is a form of induction rule. It would require some complex logical machinery to prove normalisation for rules such as these, I shall not do it here. I shall content myself to argue that the meaning of an ancestral operator is not fixed entirely by these rules. It is bold enough to claim that people use the rule $*E$ (I do claim it) but it is perhaps too bold to claim that we use them knowingly let alone as a definition. The meaning of the ancestral operator is fixed by or experiences which help us develop a concept of an ancestor or something that can be reached by repeatedly carrying out some operation. I claim that if we experience enough domino effects we develop a concept of a domino effect and thereby come to understand an ancestral operator $*$ on a binary relation. We can see that if the first domino is knocked down and the dominos are placed appropriately that all dominos will be knocked down, regardless of how many there are. We see this, I claim, mostly by our experience of domino effects.

Because we do not understand the ancestral operator by definition but mainly by a relation to our experience of structural properties of the world there is no question of a non-standard interpretation of our ancestral connective. The ancestral operator should be interpreted as expressing the *real* ancestral relation that exists between objects in the world. More precisely, we fix the meaning of the ancestral operator by our experience of objects entering into domino effects, such objects always make up standard models for the ancestral operator, thus our ancestral operator should always receive a standard interpretation.[9]

[9] A standard model for the ancestral operator is one where if aR^*b then there really is a finite sequence $c_1 \dots c_n$ such that aRc_1 and c_iRc_{i+1} $(1 \leq i < i)$ and c_nRb. We can find models that validate the rules for $*$ above but where this is not the case, such models are non-standard.

A.2.5 Plural predication

We can use the Geach construction to express plural predications. The idea is this: a sentence like 'some people lifted a piano' may be paraphrased as 'some people lifted a piano together with only each other', where '. . . lifted. . . together with. . . ' is a three place predicate. It just so happens that we can express that particular Geach construction in first order logic (the trick is in the logic of 'together with'.

What I shall do is enhance the language so that for every n-ary predicate like '. . . lifts. . . ', there is an $n+1$-ary predicate '. . . lifts. . . together with. . . ' (and also '. . . together with. . . lifts. . . ').

The language

Let there be two sorts of predicate, *simple predicates* and *complex predicates*.[10] Every complex predicate is associated with (or derives from) a simple predicate.

Let us use $F, F_1, F_2 \ldots, G, G_1 \ldots$ to denote *simple* predicates. Every complex predicate is denoted by $F^{i_1 \cdots i_m}$ where F denotes an n-ary simple predicate, and $1 \leq i_j \leq n$, and $i_1 < i_2 \cdots < i_m$, and where each i_j is a positive integer.

The complex predicate symbols $F^{i_1 \cdots i_m}$, will not be used to construct any formula of the language so they need not be given and arity (i.e. we need to say whether they are n-ary for any n).

Any variable is a term, any constant is a term. Call a variable or a constant a *simple term.* If if t_1 and t_2 are *simple terms* then $f(t_1, t_2)$ is a term, call it a *complex term.* Notice that if $f(t_1, t_2)$ is a term then neither t_1 nor t_2 contain f themselves.

- If F is an n-ary simple predicate and none of $t_1 \ldots t_n$ contain f (i.e. they are all variables and constants), then

  $$Ft_1 \ldots t_n$$

  is a *simple atomic formula.*

- If F is an n-ary simple predicate and some of $t_1 \ldots t_n$ contain f, then

  $$Ft_1 \ldots t_n$$

  is a *complex atomic formula.*

[10] Semantically, both simple and complex predicates are assigned sets of n-tuples of elements, though there are extra conditions on the set of n-tuples that can be assigned to a complex predicate.

The intuitive interpretation

If F unary then Ft should be read as:

> t is F

but $Ff(t_1, t_2)$ should be read as

> t_1 and t_2 together (i.e. collectively), possibly among others, are F

If F is binary then we may form three types of complex atomic formula using simple terms t_1, t_2, t_3, t_4: $Ff(t_1, t_2)t_3$, $Ft_1, f(t_2, t_3)$ and $Ff(t_1, t_2)f(t_3, t_4)$. The interpretation of $Ff(t_1, t_2)t_3$ is

> t_1 and t_2 together, possibly among others, relate by F to t_3

and $Ft_1 f(t_2, t_3)$ is

> t_1 relates by F to t_2 and t_3 together, possibly among others

and $Ff(t_1, t_2)f(t_3, t_4)$ is

> t_1 and t_2 together relate by F to t_3 and t_4 together.

Here are examples of how F relates to $F^{i_1 \cdots i_m}$ for the case where F is binary.

- if F is '...is walking', then $Ff(\dots, \dots)$ is '...is walking possibly among others together with... (this is also how to interpret F^1)
- if F is '...lifts...' then
  1. F^1 is '...together with...lifts...'
  2. F^2 is '...lifts...together with...' weighs...'
  3. $F^{1,2}$ is '...together with...lifts...together with...'

It should be apparent that $F^{i_1 \cdots i_m}$ is a more abstract representation of F where the i_j-th arguments of F are complex terms.

The construction 'together with' allows us, in a particular case, to use a collective property. If we wish to say that John lifted two stones (together, possibly among others) we may express this as:

> There are at least two stones x and y and John weighed x together with y.

We can use the Geach sentence to express collective properties like (C) above as follows:

'Some people are walking' is

> some people are walking together with only each other.

and 'John weighs some stones' is

> 'some stones are weighs by John together with only each other'.

But, actually, we can express this in first order logic. For 'together with, possibly among others' is transitive, symmetric and reflexive. For example, if Peter walks together with Jane and Jane is together with Jack, then Peter is together with Jack. So if F is n-ary then $F^{i_1 \cdots i_m}$ is transitive, reflexive and symmetric with respect to the arguments that are placed within the square brackets.

The formal interpretation

I shall now sketch and explain the intended semantics for the language above.

- Each n-ary simple predicate F is given the interpretation $|F|$ of a set of n-tuples of elements of a certain domain.
- Each $n+m$-ary complex predicate $F^{i_1 \cdots i_m}$ is given the interpretation $|F^{i_1 \cdots i_m}|$ of a set of n-tuples such that

  > if $z \in |F^{i_1 \cdots i_m}|$ then the i_j-th members of z are set of elements of a certain domain D, and the other members of z are merely elements of D.

  So for example, if F is ternary then $|F^{1,3}|$ is a set of n-tuples the first and third members of which are sets (of elements of D) but the second members of which is an element of D.
- If v is a valuation and t is a simple term then $v(t)$ is an element of the domain (all valuations agree on $v(t)$ if t is a constant).
- If v is a valuation and t is a complex term $f(t_1, t_2)$ then $v(t)$ is the set $\{v(t_1), v(t_2)\}$ (note that t_1 and t_2 are simple terms)
- If $Ft_1 \ldots t_n$ is a simple atomic formula then

$$v(Ft_1 \dots t_n) = T \text{ iff } \langle v(t_1) \dots v(t_n) \rangle \in |F|.$$

- If $Ft_1 \dots t_n$ is a complex atomic formula where $t_{i_1} \dots t_{i_m}$ are all the t_i (of $t_1 \dots t_n$) that are of the form $f(t, t')$ then

  $v(Ft_1 \dots t_n) = T$ iff there is an n-tuple z s.t. $z \in |F^{i_1 \dots i_m}|$ and z is identical to $\langle v(t_1) \dots v(t_n) \rangle$ except that the i_j-th members of z are supersets of the i_j-th members of $\langle v(t_1) \dots v(t_n) \rangle$.[11]

Notice that in this interpretation $f(t, t')$ is not interpreted as an object, it is interpreted in terms of a more complex predication. For example, if F is binary then $Ftf(t', t'')$ expresses what $F^2tt't''$ would express if we allowed it to be a formula of the language.

Thus we have what looks like a term f for a plural objects, but we need not interpret it as a referring expression. We interpret it as extending the predicate of which it is an argument to a more complex predicate (the predicate extended by '...together with...'. So, for example, we can get by without interpreting the use of 'and' in 'Russell and Whitehead' as forming a plural term that refers to some plural object composed of Russell and Whitehead.

A.2.6 Rules for plural predication

In the presentation of the following rules let F be an n-ary simple predicate symbol.

We must add these rules:

Symmetry

$$\frac{Ft_1 \dots t_n[t/f(t'', t')]}{Ft_1 \dots t_n[t_i/f(t', t'')]} \text{ where } t_i \text{ is a } \textit{simple} \text{ term.}$$

For example, suppose F is binary, then $Ff(a, b)x \vdash Ff(b, a)x$.

Reflexivity (introduction)

$$\frac{Ft_1 \dots t_n}{Ft_1 \dots t_n[t_i/f(t_i, t_i)]} \text{ where } t_i \text{ is a } \textit{simple} \text{ term.}$$

For example, suppose F is binary, then $Fab \vdash Faf(b, b)$, and $Faf(b, c) \vdash Ff(a, a)f(b, c)$.

[11] i.e. the members of $\langle v(t_1) \dots v(t_n) \rangle$ that are sets are subsets, though not necessarily proper subsets, of the members of z that are sets. The members of z and $\langle v(t_1) \dots v(t_n) \rangle$ that are sets occupy the the same positions in the orderings of z and $\langle v(t_1) \dots v(t_n) \rangle$.

Transitivity

$$\frac{Ft_1 \ldots t_n[t_i/f(t,t')] \quad Ft_1 \ldots t_n[t_i/f(t',t'')]}{Ft_1 \ldots t_n[t_i/f(t,t'')]}$$ where t_i is a *simple* term.

As an example suppose F is binary, then $Ff(a,b)x, Ff(b,c)x \vdash Ff(a,c)x$.

Elimination

A further rule is this:

$$\frac{Ft_1 \ldots t_n[t_i/f(t,t')] \qquad \begin{array}{c}\cancel{Ft_1 \ldots t_n[t_i/f(t,c)]}\\ \vdots \\ t = c\end{array}}{Ft_1 \ldots t_n[t_i/t]}$$

where c is a constant that does not occur in any formulae or assumptive rule applications on which $t = c$ depends, except $Ft_1 \ldots t_n[t_i/f(t,c)]$, and c does not occur in $Ft_1 \ldots t_n[t_i/f(t,t')]$. Also where t_i is a *simple* term.

as example of this rule in action, suppose F is binary and we can deduce that $Ff(aa)f(bd)$ and that $Ff(ac)f(bd) \vdash c = a$ (where c meets the restriction) then we may conclude that $Faf(bd)$. [12] To see that this rule is sound suppose that John wrote a book together with, among others, himself, and suppose further that anyone who wrote the book with John is in fact John. In other words, suppose John wrote the book alone. Then, we may say simple that John wrote a book.

[12] Those who do not like the use of $t_1 = c$ in the inferential premise may use this rule instead:

$$\frac{Ft_1 \ldots t_n[t_i/f(t,t')] \qquad \begin{array}{c}\cancel{Ft_1 \ldots t_n[t_i/f(t,c)]}, \cancel{Xt_i}\\ \vdots \\ Xc\end{array}}{Ft_1 \ldots t_n[t_i/t]}$$

with the same condition on c and where X is a simple atomic predicate that occurs in no assumptions or weak rule applications on which Xc depends (except Xt_i) and is not F, and no application of the plural predicate rules above are applied to X in the deduction of Xc. Similarly we could replace the introduction rule for $=$ with a rule like:

$$\begin{array}{c}\cancel{Xt_2}\\ \vdots \\ \dfrac{Xt_1}{t_1 = t_2}\end{array}$$

with a similar restriction on X. Given this restriction, the deduction of Xt_1 does not depend on anything about X and so X may be replace throughout by A to obtain a deduction of $A[x/t_1]$ from $A[x/t_2]$ which legitimates the usual elimination rule for $=$.

If B is a binary predicate then the binary the binary relation R such that $R(x,y) \leftrightarrow Baf(x,y)$ is transitive and hence is identical to its transitive closure (identical to its ancestral relation). So sentences like (C) are first order definable after all:

> There is a stone x and there is a stone y and John lifts x together with y, and for any z if John lifts x together with z then z a stone.

is a first order analysis of (C), more formally let L be '... lifts...' and S be '... is a stone' and let a be John:

$$\exists x \exists y[Sx \wedge Sy \wedge Laf(xy) \wedge \forall z(Laf(xz) \supset Sz)]$$

Using this extra construction we avoid explicitly predicating over collections, plural predication becomes predicating with a plural predicate rather than predicating over a plural entity. Actually, we can formalise 'John lifts some stones' in a simpler way:

$$\exists x[Sx \wedge Laf(xx) \wedge \forall z(Laf(xz) \supset Sz)]$$

and so I claim that the construction 'some stones ...' is sometimes $\exists x(Sx \wedge \ldots x \ldots)$ and, in plural cases, it is $\exists x(Sx \wedge \ldots f(xx) \ldots)$.

Therein lies my analysis of plural predication. We do not require a second order language or second order quantification in order to obtain plural predication, first order logic can handle it if we enrich its language of predicates. Of course, second order logic is still required to express more complex sentences about pluralities (but it is not required for plural predication).

Here are some more examples:

> John and Mary are moving a piano

let a and b be John and Mary respectively, let M be '... is moving...', and let P be '... is a piano':

$$\exists x(Px \wedge \forall y \forall z(Mf(yx)z \equiv ((x = a \vee x = b) \wedge (y = a \vee y = b))))$$

> John and Mary and Peter are moving a piano

let c be Peter, the formalisation is:

$$\exists x(Px \wedge \forall y \forall z(Mf(yx)z \equiv ((x = a \vee x = b \vee x = c) \wedge (y = a \vee y = b \vee y = c))))$$

> Some critics are moving a piano

$$\exists x \exists \alpha \forall y (Px \wedge \alpha(y) \supset (Cy \wedge \forall z (Mf(yx)z \equiv \alpha(z))))$$

which is the formalisation of 'there is a piano that some critics are moving together with only each other'. We can formalise the sentence without using any second order quantifiers:

$$\exists x \exists z (Px \wedge Cz \wedge Mf(zz)x \wedge \forall y (Mf(zy)x \supset Cy))$$

With this analysis we should be careful always to use present tense predicates and a tensed modal operator rather than tensed predicates. So 'John and Mary moved a piano' should be analysed as 'in the past (John and Mary are lifting a piano)'.